Proceedings of the International Congress of Mathematicians
August 16-24 1983
Warszawa

Proceedings of the International Congress of Mathematicians August 16-24 1983 Warszawa

Volume 1

PWN – Polish Scientific Publishers
Warszawa
North - Holland
Amsterdam • New York • Oxford
1984

Editors: Zbigniew Ciesielski and Czesław Olech

Cover design by Zygmunt Ziemka

Conference sign: Stefan Nargiełło

Library of Congress Cataloging in Publication Data

International Congress of Mathematicians (1983: Warsaw, Poland)
Proceedings of the International Congress of Mathematicians, August 16–24, 1983, Warszawa.

Papers in English and Russian.
Includes index.
1. Mathematics – Congresses. I. Olech, Czesław.
II. Ciesielski, Zbigniew, 1934– . III. Title.
QA1.I82 1983 510 84-13788
ISBN 0-444-86659-0 (v. 1)
ISBN 0-444-86660-4 (v. 2)
ISBN 0-444-86661-2 (set)

Published by
PWN – POLISH SCIENTIFIC PUBLISHERS – Warszawa
and
ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991, 1000 BZ Amsterdam
The Netherlands

Sole distributors for
the USA and Canada
ELSEVIER SCIENCE PUBLISHING COMPANY, Inc.
52 Vanderbilt Avenue
New York, NY 10017
USA

Albania, Bulgaria, Cuba, Czechoslovakia, Democratic People's Republic of Korea, German Democratic Republic, Hungary, Mongolia, People's Republic of China, Poland, Romania, the USSR, Vietnam and Yugoslavia
ARS POLONA
Krakowskie Przedmieście 7, 00-068 Warszawa

Printed in Poland by WDN

Contents

Volume 1

Organization of the Congress XI
Invited addresses XV
Congress members XXIII
Opening ceremonies XLIX
Closing ceremonies LIX

THE WORK OF THE MEDALLISTS

Huzihiro Araki – The work of Alain Connes 3
C. T. C. Wall – On the work of W. Thurston 11
L. Nirenberg – The work of Yau, Shing-Tung 15
J. Schwartz – The work of Robert Endre Tarjan 21

INVITED ONE-HOUR PLENARY ADDRESSES

V. I. Arnold – Singularities of ray systems 27
P. Erdös – Extremal problems in number theory, combinatorics and geometry 51
W. H. Fleming – Optimal control of Markov processes 71
C. Hooley – Some recent advances in analytical number theory 85
Wu-chung Hsiang – Geometric applications of algebraic K-theory . . . 99
P. D. Lax – Problems solved and unsolved concerning linear and nonlinear partial differential equations 119
V. P. Maslov – Non-standard characteristics in asymptotical problems 139
B. Mazur – Modular curves and arithmetic 185
R. D. MacPherson – Global questions in the topology of singular spaces 213
A. Pełczyński – Structural theory of Banach spaces and its interplay with analysis and probability 237
D. Ruelle – Turbulent dynamical systems 271
Yum-Tong Siu – Some recent developments in complex differential geometry 287

INVITED 45-MINUTE ADDRESSES IN SECTIONS

Section 1. Mathematical logic and foundations of mathematics

G. L. Cherlin – Totally categorical structures 301
J.-Y. Girard – The Ω-rule 307
P. A. Loeb – Measure spaces in nonstandard models underlying standard stochastic processes 323

R. A. Shore — The degrees of unsolvability: the ordering of functions by relative computability 337
A. O. Slisenko — Linguistic considerations in devising effective algorithms 347
B. I. Zil'ber — The structure of models of uncountably categorial theories 359

Section 2. Algebra

R. L. Gries, Jr. — The sporadic simple groups and construction of the monster 369
M. Gromov — Infinite groups as geometric objects 385
J. C. Jantzen — Einhüllende Algebren halbeinfacher Lie-Algebren . . . 393
A. Joseph — Primitive ideals in enveloping algebras 403
A. Yu. Ol'šanskiĭ — On a geometric method in the combinatorial group theory 415
C. M. Ringel — Indecomposable representations of finite-dimensional algebras 425
C. Soulé — K-théorie et zéros aux points entiers de fonctions zêta 437
R. P. Stanley — Combinatorial applications of the hard Lefschetz theorem 447
E. I. Zel'manov — On the theory of Jordan algebras 455

Section 3. Number theory

A. N. Andrianov — Integral respresentations of quadratic forms by quadratic forms: multiplicative properties 465
J.-M. Fontaine — Représentations p-adiques 475
D. R. Heath-Brown — Finding primes by sieve methods 487
D. W. Masser — Zero estimates on group varieties 493
K. A. Ribet — Congruence relations between modular forms 503
W. M. Schmidt — Analytic methods for congruences, diophantine equations and approximations 515
J.-L. Waldspurger — Correspondances de Shimura 525

Section 4. Geometry

S. Y. Cheng — On the real and complex Monge-Ampère equation and its geometric applications 533
N. J. Hitchin — The geometry of monopoles 541
A. G. Khovanskij — Fewnomials and Pfaff manifolds 549
W. Müller — Spectral geometry and non-compact Riemannian manifolds 565
R. M. Schoen — Minimal surfaces and positive scalar curvature 575
L. Simon — Recent developments in the theory of minimal surfaces . . . 579
K. K. Uhlenbeck — Variational problems for gauge fields 585
E. B. Vinberg — Discrete reflection groups in Lobachevsky spaces . . . 593
О. Я. Виро — Успехи последних 5 лет в топологии вещественных алгебраических многообразий 603

Section 5. Topology

F. R. Cohen — Applications of loop spaces to classical homotopy theory 621
R. L. Cohen — The homotopy theory of immersions 627
S. K. Donaldson — Gauge theory and topology 641
M. H. Freedman — The disk theorem for four-dimensional manifolds . . . 647

S. P. Kerckhoff — The geometry of Teichmüller space 665
Wen-Hsiung Lin — Some remarks on the Kervaire invariant conjecture . 679
J. L. Shaneson — Linear algebra, topology and number theory 685

Section 6. Algebraic geometry

A. Beilinson — Localization of representations of reductive Lie algebras 699
W. Fulton — Some aspects of positivity in algebraic geometry 711
J. Harris — Recent work on $\mathcal{M}_g$ 719
S. Iitaka — Birational geometry of algebraic varieties 727
V. A. Iskovskih — Algebraic threefolds with special regard to the problem of rationality 733
S. Mori — Cone of curves, and Fano 3-folds 747
A. Ogus — Periods of integrals in characteristic p 753
B. Teissier — Sur la classification des singularites des espaces analytiques complexes 763

Section 7. Complex analysis

W. Barth — Report on vector bundles 783
J. E. Fornæss — Holomorphic mappings between pseudoconvex domains 791
R. Harvey — Calibrated geometries 797
G. M. Henkin — Tangent Cauchy–Riemann equations and the Yang–Mills, Higgs and Dirac fields 809
P. W. Jones — Recent advances in the theory of Hardy spaces 829
С. И. Пинчук — Аналитическое продолжение отображений и задачи голоморфной эквивалентности в C^n 839

Index 847

Volume 2

Section 8. Lie groups and representations

J. Arthur — The trace formula for noncompact quotient 849
P. C. Исмагилов — Бесконечномерные группы и их представления . . 861
G. Lusztig — Characters of reductive groups over finite fields 877
P. van Moerbeke — Algebraic complete integrability of hamiltonian systems and Kac–Moody Lie algebras 881
T. Oshima — Discrete series for semisimple symmetric spaces 901
R. Parthasarathy — Unitary modules with non-vanishing relative Lie algebra cohomology 905
A. B. Venkov — The spectral theory of automorphic functions for Fuchsian groups of the first kind and its applications to some classical problems of the monodromy theory 909
M. Vergne — Formule de Kirilov et indice de l'opérateur de Dirac 921

Section 9. Real and functional analysis

R. Askey — Orthogonal polynomials and some definite integrals 935
J. Bourgain — New Banach space properties of certain spaces of analytic functions 945
B. E. J. Dahlberg — Real analysis and potential theory 953
T. Figiel — Local theory of Banach spaces and some operator ideals . . . 961
Б. С. Кашин — Некоторые результаты об оценках поперечников 977
G. G. Kasparov — Operator K-theory and its applications: elliptic operators, group representations, higher signatures, C^*-extensions 987
Y. Meyer — Intégrales singulières, opérateurs multilinéares, analyse complexe et équations aux derivées partielles 1001
B. S. Pavlov — Spectral theory of nonselfadjoint differential operators . . 1011
G. Pisier — Finite rank projections on Banach spaces and a conjecture of Grothendieck 1027
D. Voiculescu — Hilbert space operators modulo normed ideals 1041

Section 10. Probability and mathematical statistics

D. R. Brillinger — Statistical inference for random processes 1049
D. M. Chibisov — Asymptotic expansions and deficiences of tests 1063
H. Kesten — Percolation theory and resistance of random electrical networks 1081
P. Malliavin — Analyse differentielle sur l'espace de Wiener 1089
P. Mandl — Self-optimizing control of Markov processes and Markov potential theory 1097
D. W. Stroock — Stochastic analysis and regularity properties of certain partial differential operators 1107
S. Watanabe — Excursion point processes and diffusions 1117

Section 11. Partial differential equations

A. Ambrosetti — Existence and multiplicity results for some classes of nonlinear problems 1125
J.-M. Bony — Propagation et interaction des singularités pour les solutions des équations aux derivées partielles non-linéaires 1133
V. S. Buslaev — Regularization of many-particle scattering 1149
L. A. Caffarelli — Variational problems with free boundaries 1161
G. Eskin — Initial-boundary value problems for hyperbolic equations . . . 1165
E. De Giorgi — G-operators and Γ-convergence 1175
T. Iwaniec — Some aspects of partial differential equations and quasi-regular mappings 1193
S. Klainerman — Long time behavior of solutions to non-linear wave equations 1209
A. Majda — Systems of conservation laws in several space variables . . . 1217
V. E. Zakharov — Multidimensional integrable systems 1225

Section 12. Ordinary differential equations and dynamical systems

A. Katok — Nonuniform hyperbolicity and structure of smooth dynamical systems 1245

A. Lasota — Asymptotic behaviour of solutions: statistical stability and chaos . . . 1255
R. Mañé — Osceledec's theorem from the generic viewpoint . . . 1269
M. Misiurewicz — One-dimensional dynamical systems . . . 1277
G. R. Sell — Linearization and global dynamics . . . 1283

Section 13. Mathematical physics and mechanics

M. Aizenman — Stochastic geometry in statistical mechanics and quantum field theory . . . 1297
J. M. Ball — Energy-minimizing configurations in nonlinear elasticity . . . 1309
O. Ladyženskaya — On finding symmetrical solutions of field theories variational problems . . . 1315
L. A. Takhtajan — Integrable models in classical and quantum field theory . . . 1331
S. Woronowicz — Duality in the C^*-algebra theory . . . 1347

Section 14. Control theory and optimization

R. W. Brockett — Nonlinear control theory and differential geometry . . . 1357
H. W. Knobloch — Nonlinear systems: local controllability and higher order necessary conditions for optimal solutions . . . 1369
A. B. Kuržanskiĭ — Evolution equations for problems of control and estimation of uncertain systems . . . 1381
P. L. Lions — Hamilton–Jacobi–Bellman equations and the optimal control of stochastic systems . . . 1403
R. T. Rockafellar — Differentiability properties of the minimum value in an optimization problem depending on parameters . . . 1419
J. Zabczyk — Stopping problems in stochastic control . . . 1425

Section 15. Numerical methods

Feng Kang — Finite element method and natural boundary reduction 1439
R. Glowinski — Numerical solution of nonlinear boundary value problems by variational methods. Applications . . . 1455
Yu. A. Kuznetsov — Matrix iterative methods in subspaces . . . 1509
C. A. Micchelli — Recent progress in multivariate splines . . . 1523
M. J. D. Powell — On the rate of convergence of variable metric algorithms for unconstrained optimization . . . 1525

Section 16. Combinatorics and mathematical programming

D. Foata — Combinatoire des identités sur les polynômes orthogonaux . . . 1541
R. L. Graham — Recent developments in Ramsey theory . . . 1555
L. G. Khachiyan — Convexity and complexity in polynomial programming 1569
J. H. van Lint — Partial geometries . . . 1579
L. Lovász — Algorithmic aspects of combinatorics, geometry and number theory . . . 1591

Section 17. Computer and information sciences

R. Karp — The probabilistic analysis of combinatorial algorithms . . . 1601
A. A. Letichevsky — Abstract data types and finding invariants of programs 1611

R. E. Tarjan — Efficient algorithms for network optimization 1619
L. G. Valiant — An algebraic approach to computational complexity . . . 1637

Section 18. New applications of mathematics

N. Kopell — Forced and coupled oscillators in biological applications . . . 1645
B. B. Mandelbrot — On fractal geometry, and a few of the mathematical questions it has raised 1661
Yu. M. Svirezhev — Modern problems of mathematical ecology 1677

Section 19. History and education

H. Freudenthal — The implicit philosophy of mathematics history and education 1695
А. Б. Погорелов — О преподавании геометрии в школе 1711
J. B. Serrin — The structure and laws of thermodynamics 1717

Index 1729

Organization of the Congress

The recommendation to hold the 1982 International Congress of Mathematicians in Warsaw was made by the Site Committee of the International Mathematical Union during the Helsinki Congress in August 1978. The final decision was taken a few days later when the Congress, at its closing session, accepted the invitation to Warsaw, made by Professor Kazimierz Urbanik on behalf of the Polish National Committee for Mathematics.

At the beginning of 1982, due to the events in Poland, the question of holding the Congress in Warsaw was raised and discussed again. In April 1982, the Executive Committee of IMU, considering the scientific prospects for ICM-82 at that time, decided to postpone the Warsaw Congress by one year. At the same time the Executive Committee decided to hold the General Assembly of IMU in Warsaw, as previously planned, in August 1982. The question of the Warsaw Congress was extensively discussed at that meeting. Eventually, in November 1982, the Executive Committee of IMU finally confirmed the organization of ICM-82 in Warsaw in August 1983.

Poland's offer to be the host of the ICM-82 followed upon the promise of support made by the Polish Academy of Sciences. The Congress was particularly honoured by the fact that Professor Aleksander Gieysztor, President of the Polish Academy of Sciences had consented to be its Patron.

The scientific programme was the responsibility of the International Mathematical Union, acting through the Consultative Committee, whose members were Professors J. P. Serre (chairman), M. Atiyah, B. Bojarski, W. Browder, Z. Ciesielski, P. Deligne, L. Faddeev, S. Łojasiewicz and S. Winograd. The Committee was established in 1979 and in June 1980 decided to divide the mathematical programme into 19 sections and appointed the cores of the panels for those sections. The panels were finally set up and submitted their suggestions before the summer of 1981. Considering these suggestions and also suggestions received from some National Committies, the Consultative Committee in October 1981 selected 16 mathematicians to give one-hour plenary addresses and 137 to give 45-minute addresses in the sections. Some more names were added later, three after it was known that ICM-82 had been postponed. All the persons invited to deliver the plenary addresses and 129 of those asked to speak in the sections accepted the invitation. Of the total number of 145 prospective speakers 110 were present at the Congress including 13 plenary speakers. The manuscripts of 5 absent speakers were read out at the Congress.

The Fields Medals Committee consisted of Professors L. Carleson (chairman ex officio), H. Araki, N. Bogolyubov, P. Malliavin, D. Mumford, L. Nirenberg, A. Schinzel and C. T. C. Wall. The committee for the newly established Nevanlinna prize in "Mathematical aspects of Information Science" consisted of Professors J.-L. Lions (chairman), J. Schwartz and A. Salomaa. The decisions of these Committees were announced at the General Assembly meeting in August 1982.

Other preparations for the Congress were in the hands of the Polish Organizing Committee. Its chairman was Czesław Olech, who took direct responsibility for all arrangements. A great amount of work was done by Bogdan Bojarski, Jerzy Browkin, Zbigniew Ciesielski, Eugeniusz Fidelis, Stanisław Łojasiewicz, Zbigniew Semadeni, Wiesław Żelazko and other members of the committee. Altogether more than 60 Polish mathematicians took part in the preparations for the Congress. A small Congress Bureau was set up in 1979, assisted by the administrative staff of the Institute of Mathematics of the Polish Academy of Sciences. A particularly important role was played by Anna Sierpińska-Jankowska, who was fully engaged in the affairs of the Congress from the very beginning. Registration, accommodation and some other arrangements were handled by the Orbis Congress Bureau.

The main sources of funds for the Congress were:

(1) a subvention from the Polish Academy of Sciences,

(2) a subvention from the International Mathematical Union,

(3) membership fees.

All financial matters were attended to by the Institute of Mathematics of the Polish Academy of Sciences. Some facilities were offered to the Congress without charge by the Warsaw University.

The International Mathematical Union gave travel grants to young mathematicians from developing countries and the Congress waived their fees.

A short preliminary announcement about the Warsaw Congress was sent out in the autumn of 1980 to all the countries of the world in which mathematical communities were known to exist. The First Announcement was dispatched in July 1981 to the same addresses with the request that copies of it be further distributed among the mathematicians of the countries in question. The Second Announcement, containing detailed information about the Congress, planned for August 1982, and including the registration form, was mailed in December 1981 to those mathematicians who had applied for it; about 6000 copies were sent, mostly to individual addresses. Only a small number of the registration forms were returned in the spring of 1982.

The information about the April 1982 decision of the Executive Committee of IMU was mailed in May 1982 to the same addresses. When it was finally decided to hold the Congress in Warsaw in 1983, the Third Announcement was issued. It contained all the information given in the Second Announcement brought up to date and also a list of the invited speakers with the titles of their addresses as well as a rough schedule. The mailing of this announcement started towards the end of January 1983.

2400 ordinary members of the Congress were registered and about 150 accompanying persons from more than 60 countries; not all of them eventually turned up. The lectures and seminars were also attended by a number of non-registered participants. The mathematical activities of the Congress took place in the Palace of Culture, located in the centre of Warsaw.

Besides the invited addresses included in the official programme, about 680 short communications were presented during the Congress. Summaries of more than 800 short communications which reached the organizers in time were photocopied and offered to the Congress members upon registration. A symposium of the International Commission on Mathematical Instruction was held on four afternoons, accompanied by a seminar of the International Group on the Relations between History and Pedagogy of Mathematics affiliated to ICMI. On the initiative of Congress members 14 different seminars were organized, some of them having more than one session. A show of computer-animated films was presented on two evenings by the author Professor T. F. Banchoff.

A book exhibition organized in cooperation with the ORPAN — Distribution Centre of Scientific Publications of the Polish Academy of Sciences was open throughout the Congress.

The City of Warsaw showed its hospitality to the participants by granting them free travel on all city buses and trams for the duration of the Congress.

On the evening of August 19 the President of the Polish Academy of Sciences gave a reception to which all Congress members and the accompanying persons were invited.

The Organizing Committee arranged various social events. On Saturday August 20 and Sunday August 21 excursions to Bogusławice were organized. Each included a picnic party and a show called "The Cracovian wedding". The latter was a spectacle comprising folk songs and dances in the colourful dresses of the Cracow region and a parade of riders and equipages driven by horses in the traditional Cracovian harnesses. The excursions had about 1200 participants each. On the same days, for those members and guests of the Congress who remained in Warsaw, the Silesian folklore ensemble "Śląsk" gave a special performance. Two violin recitals by Aureli Błaszczok (violin) and Maria Szwajger-Kułakowska (piano accompaniment) were given on August 17 and August 18.

Invited Addresses

Most of the speakers who accepted the invitation gave their addresses in the Congress themselves and submitted manuscripts for printing. If this was not the case a number (1), (2), (3) or (4) appears after the speaker's name in the list below. These numbers have the following meaning:

(1) The speaker did not attend the Congress. His manuscript was read there, and it is printed in the Proceedings.

(2) The speaker did not attend the Congress. His lecture was cancelled, but his manuscript is printed in the Proceedings.

(3) The speaker delivered the address in the Congress but did not submit a manuscript for the Proceedings.

(4) The speaker did not attend the Congress and did not send a manuscript.

One-hour addresses

V. I. Arnold — Singularities of ray systems

P. Erdös — Extremal problems in number theory, combinatorics and geometry

W. H. Fleming — Optimal control of Markov processes

C. Hooley — Some recent advances in analytical number theory

Wu-chung Hsiang — Geometric applications of algebraic K-theory

P. D. Lax — Problems solved and unsolved concerning linear and nonlinear partial differential equations

V. P. Maslov — Non-standard characteristics in asymptotical problems

B. Mazur — Modular curves and arithmetic

R. D. MacPherson — Global questions in the topology of singular spaces

A. Pełczyński — Structural theory of Banach spaces and its interplay with analysis and probability

M. Rabin (4) — Computational complexity and randomizing algorithms

D. Ruelle — Turbulent dynamical systems

M. Sato (3) — Monodromy theory and holonomic quantum fields — a new link between mathematics and theoretical physics

S. Shelah (4) — On some problems on the continuum

Yum-Tong Siu — Some recent developments in complex differential Geometry
R. Thom (4) — Mathematics and scientific explanation

45-minute addresses in sections

Section 1. Mathematical logic and foundations of mathematics

G. L. Cherlin — Totally categorical structures
J.-Y. Girard — The Ω-rule
P. A. Loeb — Measure spaces in nonstandard models underlying standard stochastic processes
R. A. Shore — The degrees of unsolvability: the ordering of functions by relative computability
A. O. Slisenko — Linguistic considerations in devising effective algorithms
B. I. Zil'ber — The structure of models of uncountably categorical theories

Section 2. Algebra

R. L. Griess, Jr. — The sporadic simple groups and construction of the monster
M. Gromov (2) — Infinite groups as geometric objects
J. C. Jantzen — Einhüllende Algebren halbeinfacher Lie-Algebren
A. Joseph (2) — Primitive ideals in enveloping algebras
A. Yu. Ol'šanskiĭ — On a geometric method in the combinatorial group theory
C. M. Ringel — Indecomposable representations of finite-dimensional algebras
C. Soulé — K-théorie et zéros aux points entiers de fonctions zêta
R. P. Stanley — Combinatorial applications of the hard Lefschetz theorem
E. I. Zel'manov — On the theory of Jordan algebras

Section 3. Number theory

A. N. Andrianov — Integral representations of quadratic forms by quadratic forms: multiplicative properties
J.-M. Fontaine — Représentations p-adiques
D. R. Heath-Brown — Finding primes by sieve methods
D. W. Masser — Zero estimates on group varieties
K. A. Ribet — Congruence relations between modular forms

W. M. Schmidt (2) — Analytic methods for congruences, diophantine equations and approximations
J.-L. Waldspurger — Correspondances de Shimura

Section 4. Geometry

S. Y. Cheng — On the real and complex Monge–Ampère equation and its geometric applications
N. J. Hitchin — The geometry of monopoles
A. G. Khovanskiĭ — Fewnomials and Pfaff manifolds
W. Müller — Spectral geometry and non-compact Riemannian manifolds
R. M. Schoen (2) — Minimal surfaces and positive scalar curvature
L. Simon — Recent developments in the theory of minimal surfaces
K. K. Uhlenbeck (2) — Variational problems for gauge fields
E. B. Vinberg (1) — Discrete reflection groups in Lobachevsky spaces
О. Я. Виро (1) — Успехи последних 5 лет в топологии вещественных алгебраических многообразий

Section 5. Topology

F. R. Cohen — Applications of loop spaces to classical homotopy theory
R. L. Cohen (2) — The homotopy theory of immersions
S. K. Donaldson — Gauge theory and topology
M. H. Freedman (2) — The disk theorem for four-dimensional manifolds
S. P. Kerckhoff (2) — The geometry of Teichmüller space
Wen-Hsiung Lin — Some remarks on the Kervaire invariant conjecture
J. L. Shaneson (2) — Linear algebra, topology and number theory
H. Toruńczyk (3) — On the topology of infinite-dimensional manifolds

Section 6. Algebraic geometry

A. Beilinson — Localization of representations of reductive Lie algebras
W. Fulton (2) — Some aspects of positivity in algebraic geometry
J. Harris (2) — Recent work on $\mathcal{M}_g$
S. Iitaka — Birational geometry of algebraic varieties
V. A. Iskovskih — Algebraic threefolds with special regard to the problem of rationality
S. Mori — Cone of curves, and Fano 3-folds
A. Ogus — Periods of integrals in characteristic p
B. Teissier — Sur la classification des singularites des espaces analytiques complexes

Section 7. Complex analysis

W. Barth — Report on vector bundles
J. E. Fornaess (2) — Holomorphic mappings between pseudoconvex domains
R. Harvey (2) — Calibrated geometries
G. M. Henkin — Tangent Cauchy–Riemann equations and the Yang–Mills, Higgs and Dirac fields
P. W. Jones — Recent advances in the theory of Hardy spaces
С. И. Пинчук — Аналитическое продолжение отображений и задачи голоморфной эквивалентности в C^n

Section 8. Lie groups and representations

J. Arthur (2) — The trace formula for noncompact quotient
Р. С. Исмагилов — Бесконечномерные группы и их представления
G. Lusztig — Characters of reductive groups over finite fields
P. van Moerbeke — Algebraic complete integrability of hamiltonian systems and Kac–Moody Lie algebras
T. Oshima — Discrete series for semisimple symmetric spaces
R. Parthasarathy — Unitary modules with non-vanishing relative Lie algebra cohomology
A. B. Venkov — The spectral theory of automorphic functions for Fuchsian groups of the first kind and its applications to some classical problems of the monodromy theory
M. Vergne — Formule de Kirilov et indice de l'opérateur de Dirac

Section 9. Real and functional analysis

R. Askey — Orthogonal polynomials and some definite integrals
J. Bourgain — New Banach space properties of certain spaces of analytic functions
B. E. J. Dahlberg (2) — Real analysis and potential theory
T. Figiel — Local theory of Banach spaces and some operator ideals
Б. С. Кашин — Некоторые результаты об оценках поперечников
G. G. Kasparov — Operator K-theory and its applications: elliptic operators, group representations, higher signatures, C^*-extensions
Y. Meyer — Intégrales singulières, opérateurs multilinéares, analyse complexe et équations aux derivées partielles
B. S. Pavlov — Spectral theory of nonselfadjoint differential operators
G. Pisier — Finite rank projections on Banach spaces and a conjecture of Grothendieck
D. Voiculescu — Hilbert space operators modulo normed ideals

Section 10. Probability and mathematical statistics

D. R. Brillinger (1) — Statistical inference for random processes
D. M. Chibisov — Asymptotic expansions and deficiences of tests
H. Kesten (2) — Percolation theory and resistance of random electrical networks
P. Malliavin — Analyse différentielle sur l'espace de Wiener
P. Mandl — Self-optimizing control of Markov processes and Markov potential theory
D. W. Stroock (2) — Stochastic analysis and regularity properties of certain partial differential operators
S. Watanabe — Excursion point processes and diffusions

Section 11. Partial differential equations

A. Ambrosetti — Existence and multiplicity results for some classes of nonlinear problems
J.-M. Bony — Propagation et interaction des singularités pour les solutions des équations aux derivees partielles non-linéaires
V. S. Buslaev — Regularization of many-particle scattering
L. A. Caffarelli (2) — Variational problems with free boundaries
G. Eskin — Initial-boundary value problems for hyperbolic equations
E. De Giorgi — G-operators and Γ-convergence
T. Iwaniec — Some aspects of partial differential equations and quasi-regular mappings
S. Klainerman — Long time behavior of solutions to nonlinear wave equations
A. Majda (2) — Systems of conservation laws in several space variables
V. E. Zakharov — Multidimensional integrable systems

Section 12. Ordinary differential equations and dynamical systems

A. Katok (1) — Nonuniform hyperbolicity and structure of smooth dynamical systems
A. Lasota — Asymptotic behaviour of solutions: statistical stability and chaos
R. Mañé — Oseledec's theorem from the generic viewpoint
M. Misiurewicz — One-dimensional dynamical systems
G. R. Sell — Linearization and global dynamics

Section 13. Mathematical physics and mechanics

M. Aizenman — Stochastic geometry in statistical mechanics and quantum field theory

J. M. Ball — Energy-minimizing configurations in nonlinear elasticity
O. Ladyženskaya — On finding symmetrical solutions of field theories variational problems
T. Nishida (4) — Equation of compressible, viscous and heatconductive fluids and equation of Boltzmann
K. Osterwalder (4) — Recent progress towards the construction of quantized fields
L. A. Takhtajan — Integrable models in calssical and quantum field theory
S. Woronowicz — Duality in the C^*-algebra theory

Section 14. Control theory and optimization

R. W. Brockett — Control theory and differential geometry
H. W. Knobloch — Nonlinear systems: local controllability and higher order necessary conditions for optimal solutions
A. B. Kuržanskiĭ — Evolution equations for problems of control and estimation of uncertain systems
P.-L. Lions — Hamilton–Jacobi–Bellman equations and the optimal control of stochastic systems
R. T. Rockafellar (2) — Differentiability properties of the minimum value in an optimization problem depending on parameters
J. Zabczyk — Stopping problems in stochastic control

Section 15. Numerical methods

B. Engquist (3) — Computational far field boundary conditions for partial differential equations
Feng Kang (2) — Finite element method and natural boundary reduction
R. Glowinski — Numerical solution of nonlinear boundary value problems by variational methods. Applications
G. H. Golub (4) — Some inverse matrix problems
Yu. A. Kuznetsov — Matrix iterative methods in subspaces
C. A. Micchelli — Recent progress in multivariate splines
M. J. D. Powell (1) — On the rate of convergence of variable metric algorithms for unconstrained optimization

Section 16. Combinatorics and mathematical programming

D. Foata — Combinatoire des identités sur les polynômes orthogonaux
R. L. Graham — Recent developments in Ramsey theory
L. G. Khachiyan (2) — Convexity and complexity in polynomial programming

Congress Members

An asterisk marks those who did not attend the Congress

ABDEL-MEGIED M., Egypt
ABE Michio, Japan
ABISH Mekei, Mongolia
ADAMOWICZ Zofia, Poland
ADAMS John Frank, UK
ADAMSON I. T., UK
ADOLFSSON K. V., Sweden
ADOLFSSON Vilhelm, Sweden
ADYAN S. I., USSR
AFGHAHI Mohammad H., Iran
AGAYAN S. S., USSR
AGOSHKOV V. I., USSR
AGRACHEV A. A., USSR
AHRENS Heinz, GDR
AIZENMAN M., USA
AKIVIS M. A., USSR
AKYILDIZ Yilmaz, Turkey
AL-AWAMI Mohamed M., Libya
ALBRYCHT Jerzy, Poland
ALEKSANDROV Aleksej B., USSR
ALEŠKEVIČENE Aldona K., USSR
ALEXIEWICZ Andrzej, Poland
ALFARO Manuel, Spain
ALFAWICKA Barbara, Poland
ALYMKULOV K., USSR
AMBROSETTI A., Italy
AMBROSIEWICZ Jan, Poland
AMEMIYA Frances C., USA
AMITSUR S. A., Israel
ANAND Kailash K., Canada
ANDERSEN Henning Haahr, Denmark
ANDERSON Robert L., USA
ANDREIAN CAZACU Cabiria, Romania
ANDRIANOV A. N., USSR
ANDRIÉ M., FRG
ANDRUNAKIEVIČ Vladimir, USSR
ANDRZEJCZAK Grzegorz, Poland
ANDRZEJEWSKI Paweł, Poland
ANISZCZYK Bohdan, Poland
ANOSOV D. V., USSR
APPENZELLER F. U., Switzerland
ARAK Tairo V., USSR
ARAKI Huzihiro, Japan
ARCAVI Abraham, Israel
ARMINJON Paul G., Canada
ARNAUTOV V. I., USSR
ARNOLD V. I., USSR
ARROWSMITH David K., UK
ARSZYŃSKA Jolanta, Poland
ARTHUR James, Canada*
ASADA Teruko, Japan
ASAM Lilian, FRG
ASIBONG-IBE U. I., Nigeria
ASKEY R. A., USA
ASSER Günter, GDR
ASTALA K. A., Finland
ATIYAH M. F., UK
AUDIN Michèle, France
AUERT Fritz, GDR
AULASKARI R. P. T., Finland
AUPETIT B., Canada
AVCI Yusuf, Turkey
AXLER Sheldon, USA

BACKELIN Jörgen, Sweden
BADALYAN G. V., USSR
BADJI Cherif, Senegal
BAJPAI S. D., Nigeria
BAKHVALOV N. S., USSR
BAKSALARY Jerzy, Poland
BALCERZYK Stanisław, Poland
BALL J. M., UK
BALOG Antal, Hungary
BALOGH Zoltán, Hungary
BAMBA Siaka Kante, Ivory Coast

BANAŚ Józef, Poland
BANASIEWICZ Bogdan, Poland
BANASZAK Grzegorz, Poland
BANCHOFF T. F., USA
BANULESCU Martha, Romania
BARAN A., Romania
BARANOWSKI Jan, Poland
BARAŃSKI Feliks, Poland
BARBU Vasile, Romania
BARCZ Eugeniusz, Poland*
BARON Karol, Poland
BARSOOM Magdy, Egypt
BARTH Wolf P., FRG
BARTOL Witold, Poland
BARTOSIEWICZ Z., Poland
BARTOSZEK W., Poland
BARTOSZEWICZ Jarosław, Poland
BARTOSZYŃSKI Tomasz, Poland
BASHKENE A. L., USSR
BASTERO Jesús, Spain
BATLE Nadal, Spain
BAUDISCH Andreas, GDR
BAUMGÄRTEL H., GDR
BECK József, Hungary
BECKER Hans J., FRG
BECKERT Herbert, GDR
BĘCZEK Jan, Poland
BEDNARCZUK Ewa, Poland
BEDNARCZUK Jerzy, Poland
BEDNARCZYK Mirosław, Poland
BEGEHR H., FRG
BÉGUERI Lucile, France
BEILINSON A. A., USSR
BEKKA B., Algeria
BEKLEMISHEVA L. A., USSR
BELEGRADEK O. V., USSR
BELL Harold, USA
BELOWSKA Laura, Poland
BENKERT Frank, GDR
BERCEANU B., Romania
BEREZANSKII Yu. M., USSR
BERG Christian, Denmark
BERKES I., Hungary
BERLING Torgöt, Sweden
BERTAUD M., Canada
BERTIN M. J., France
BERTRAND Daniel A., France
BESALA Piotr, Poland
BEŚKA Marek, Poland
BESSAGA Czesław, Poland
BEZBORODOV Vyacheslav M., USSR
BEZDEK András, Hungary
BHATIA R., India
BIAŁAS Józef, Poland
BIAŁYNICKI-BIRULA Iwo, Poland
BIELECKI Tomasz, Poland
BIELSKI Włodzimierz, Poland
BIERSKI Franciszek, Poland*
BIERSTONE Edward, Canada
BIESZK Leon, Poland
BIHARI Imre, Hungary
BILER Piotr, Poland
BIRMAN M. Sh., USSR
BITSADZE A. V., USSR
BITTNER Ryszard, Poland
BJÖRCK Göran, Sweden
BŁASZCZYK Aleksander, Poland
BŁASZCZYK-ŚWIĄTKIEWICZ Bronisława, Poland
BLATTER Christian, Switzerland
BŁAŻ Jan, Poland
BLAŽEK Jaroslav, Czechoslovakia
BLIEV Nazarbaj K., USSR
BLUMENFELD Matthias, FRG
BOBIŃSKI Zbigniew, Poland
BOBOC Nicu, Romania
BOBROWSKI Dobiesław, Poland
BOBYLEV A. V., USSR
BOCHENEK Jan, Poland
BOCHKAREV S. V., USSR
BOGACKA Barbara, Poland
BOGDAN Marek, Poland
BOGNÁR János, Hungary
BOGNÁR M., Hungary
BOGOBOWICZ Agnieszka, Poland
BOGOLYUBOV N. M., USSR
BOGOMOLOV A. M., USSR
BOGOMOLOV F. A., USSR
BOGOYAVLENSKII Oleg I., USSR
BOGUCKA-KAMIŃSKA Stanisława, Poland
BÖHM Johannes, GDR
BOILLAT J., Switzerland
BOJANOV Borislav, Bulgaria
BOJARSKI Bogdan, Poland
BOJU Valentin, Romania

BOKOR József, Hungary
BOLLEN J. A. M., Netherlands
BOMAN Jan, Sweden
BONY Jean-Michel, France
BOOSS Bernheim, Denmark
BORCEA C. S., Romania
BOREVICH Z. I., USSR
BORISOV Adrijan, Bulgaria
BÖRJESON J. L., Sweden
BORON Leo F., USA
BOROWCZYK J., France
BOROWIEC Andrzej, Poland
BOROWIECKI Mieczysław, Poland
BOROWIK Nadzieja, Poland
BOROWIK Sergiusz, Poland
BOROZDIN K. V., USSR
BORTNIK Tomasz, Poland
BORZYMOWSKI A. M., Poland
BORZYSZKOWSKI Andrzej, Poland
BOSECK H., GDR
BOTHE H. G., GDR
BOURDEAU Marc, Canada
BOUTEFNOUCHET Mahmoud, Algeria
BOURGAIN Jean, Belgium
BOWSZYC Cezary, Poland
BOYAŽHIEV Peter, Bulgaria
BOŻEJKO Marek, Poland
BOZHÜYÜK M. E., Turkey
BOŽINOV N., Bulgaria
BRANDSTÄDT Andreas, GDR
BRANDWEIN Piotr, Poland
BRIEM E., Iceland
BRINDZA Bela, Hungary
BRINKHUIS J., Netherlands
BRITO Fabiano, Brazil
BROCKETT Roger W., USA
BROUSSEAU G. P. L., France
BROWDER W., USA
BROWKIN Jerzy, Poland
BROWN Leon, USA
BRUNOVSKÝ Pavol, Czechoslovakia
BRYDAK Dobiesław, Poland
BRYŃSKI Maciej, Poland
BRZEZIŃSKI Juliusz, Sweden
BRZEŹNIAK Zdzisław, Poland
BRZUCHOWSKI Jan, Poland
BRZYCHCZY Stanisław, Poland
BUCHNER Klaus, FRG
BUCKI A. J., Poland
BUGROV Ya. S., USSR
BUIANOUCKAS F. R., USA
BUIUM A., Romania
BUJALANCE Eladio, Spain
BUKOVSKÝ L., Czechoslovakia
BULA Witold, Poland
BULDYREV V. S., USSR
BULIŃSKA Maria, Poland
BURCHULADZE T. V., USSR
BURENKOV Victor I., USSR
BUREŠ Jarolím, Czechoslovakia
BURMEISTER W., GDR
BURNAT Marek, Poland*
BURYS Stanisław, Poland
BURZYK Józef, Poland
BUSETTO Giorgio, Italy*
BUSHAW D. W., USA
BUSLAEV Viktor I., USSR
BUSLAEV V. S., USSR
BUTKIEWICZ Jan, Poland
BUTSAN George P., USSR
BUTTON L. G., UK
BUZÁSI Károly, Hungary
BYCZKOWSKI T., Poland
BYRNE C. M., FRG

CAGNAC F., Cameroon
CAIN B. E., USA
CALDWELL James, UK
CAŁKA Aleksander, Poland
CAMPBELL C. M., UK
CAMPILLO Antonio, Spain
CAMPOS L. M. B. C., UK
CANOV Vasil V., Bulgaria
CARL Bernd, GDR
CARLESON Lennart, Sweden
CARLSSON Renate, FRG
CARTIER Pierre E., France
CARVALHO A. M., Mozambique
CASADIO Giuseppina, Italy
CASSELS J. W. S., UK
CASTRO Abel, Cuba
CATANĂ Viorel, Romania
CEGIEŁKA Kazimierz, Poland
CEGŁA Wojciech, Poland
CEGRELL Urban, Sweden
CHAHAL Jasbir Singh, India

CHAKARIN Kerope, Bulgaria
CHAMBERS Ll. G., UK
CHANTURIYA Z. A., USSR
CHARATONIK Janusz J., Poland
CHAUMERET Brigitte, France
CHENG Shiu-Yuen, USA
CHENTSOV A. G., USSR
CHERLIN G. L., USA
CHERNIKOV N. S., USSR
CHERNOGOROV Venelin, Bulgaria
CHERNYATIN Valentin A., USSR
CHEVALIER Bernard, France
CHEVALLIER Benoit, France
CHIBISOV D. M., USSR
CHILLINGWORTH D. R. J., UK
CHIRKA E. M., USSR
CHMAJ Jerzy, Poland
CHMIELOWSKI Jan, Poland
CHMURA Andrzej, Poland
CHOBANOV Prodan, Bulgaria
CHOCZEWSKI Bogdan, Poland
CHOJNACKI Wojciech, Poland
CHOQUET Gustave, France
CHRIST L. E., USA
CHRISTIANSEN Bent, Denmark
CHROBAK Marek, Poland
CHRONOWSKI Antoni, Poland
CHRZAN Piotr, Poland
CHRZĘSZCZYK Andrzej, Poland
CHU C. H., UK
CHUKWU E.-N., Nigeria
CICHOCKI Tadeusz, Poland*
CICHOŃ Jacek, Poland
CIEMNOCZOŁOWSKI Jarosław, Poland
CIESIELSKI Bolesław, Poland
CIESIELSKI Krzysztof, Poland
CIESIELSKI Zbigniew, Poland
CIOSEK Marianna, Poland
CLEMENTS John C., Canada
COHEN F. R., USA
COHEN G. D., France
COHN P. M., UK
COLLINS P. J., UK
CONNES Alain, France
CONNETT W. C., USA
CONSTANTINESCU A., Romania
CORBACHO Eusebio, Spain
CORNU Bernard, France
COSTA Antonio, Spain
COSTABEL Martin, FRG
CRAČIUN C., Romania
CREE George C., Canada
CRILLY A. J., UK
CRUMEYROLLE Albert, France
CSÁSZÁR Ákos, Hungary
CSIRMAZ László, Hungary
CUPAŁA Wiesław, Poland
CZECHOWSKI Tadeusz, Poland
CZERNI Marek, Poland
CZOCHRALSKA Izabella, Poland
CZUBA Stanisław T., Poland
CZUGAŁA Józefa, Poland
CZYŻ J., Poland

DĄBEK Marian, Poland*
DĄBROWSKA Róża, Poland
DAHN B. I., GDR
DALE K. T., Norway
DALY Hope H., USA
D'AMBROSIO Ubiratan, Brazil*
DAMEK Ewa, Poland
DANECKI R., Poland
DANIELYAN E. A., USSR
DANILYUK I. I., USSR
DAŃKO Wiktor, Poland
DANZER L. W., FRG
DARGAHI-NOUBARY G. R., Iran
DARK R. S., Ireland
DASSOW J., GDR
DASZKIEWICZ Andrzej, Poland
DAVIES R. O., UK
DAVIS Chandler, Canada
DAWIDOWICZ Alina, Poland
DAWIDOWICZ Antoni L., Poland
DEACONESCU M., Romania
DEÁK Ervin, Hungary
DECHAMPS-GONDIM Myriam, France
DECIU Elena, Romania
DEGTEV A. N., USSR
DE GIORGI Ennio, Italy
DE GRAAF J., Netherlands
DEKSTER Boris, Canada
DE LA TORRE Alberto, Spain
DE LANGE J., Netherlands
DELIGNE Pierre, France
DE MARR Ralph E., USA

DEMIRSKA Ewa, Poland
DEMKOWICZ Leszek, Poland
DEMYANOV Vladimir F., USSR
DENKOVA Vesselina, Bulgaria
DENKOWSKA Zofia, Poland
DENKOWSKI Zdzisław, Poland
DENNEBERG D., FRG
DERRIENNIC Marie, France
DEVIATOV B. I., USSR
DEZA M.-M., France
DEZIN A. A., USSR
DIBLÍK Josef, Czechoslovakia
DIMCA Alexandru, Romania
DIMIEV Stancho, Bulgaria
DIMITROV Jordan, Bulgaria
DINCA Gheorghe, Romania
DINEVA Petia, Bulgaria
DINGER Ulla, Sweden
DIXMIER Suzanne, France
DJOKOVIĆ D. Ž., Canada
DLAB V., Canada
DŁOTKO Tadeusz, Poland
DŁUGOSZ Jolanta, Poland
DMITRIEV M. G., USSR
DOBRAKOV Ivan, Czechoslovakia
DOBREŃKO Roman, Poland
DOBROWOLSKI Edward, Poland
DODZIUK Józef, USA
DO HONG TAN, Viet-Nam
DOITCHINOV D., Bulgaria
DOLANSKÝ Petr, Czechoslovakia
DOLBEAULT P., France
DOLBILIN N. R., USSR
DOLEŽAL Jaroslav, Czechoslovakia
DOLEŽAL Vladimir, Czechoslovakia
DOLIŃSKI Ryszard, Poland
DOMAN Ryszard, Poland
DOMAŃSKI Włodzimierz, Poland
DOMINGUEZ Eladio, Spain
DOMINGUEZ J. M., Spain
DOMORADZKI Stanisław, Poland
DONALDSON S. K., UK
DONG Guang-Chang, China*
DOROSLAVAČKI Rade, Yugoslavia
DOWBOR Piotr, Poland
DOWNAROWICZ Małgorzata, Poland
DRÁBEK Pavel, Czechoslovakia
DRENSKIĬ Vesselin, Bulgaria
DREWNIAK Józef, Poland
DREWNOWSKI Lech, Poland
DROGEANU Ioana, Romania
DROZDOWSKI Grzegorz, Poland
DROZHZHINOV Yu. N., USSR
DRUŻKOWSKI Ludwik, Poland
DUDA Roman, Poland
DUDEK Józef, Poland
DUMA Ewa, Poland
DUNCAN Bożenna, Poland
DUNCAN R., USA
DUNCAN Tyrone E., USA
DURFEE A. H., USA
DUSZAK Zbigniew, Poland
DWILEWICZ Roman, Poland
DYBEL Halina, Poland
DYMYTROV Aleksander, Bulgaria
DZEDZEJ Zdzisław, Poland
DZHAKOV Plamen B., Bulgaria
DZHRBASHYAN M. M., USSR
DZHURAEV A., USSR
DZHURAEV T. D., USSR
DZIECHCIŃSKA-HALAMODA Zyta, Poland
DZINOTYIWEYI Heneri A. M., Zimbabwe
DZIURZYŃSKI Ryszard, Poland
DŻURA Bolesław, Poland
DZYADYK V. K., USSR

ECKSTEIN F., FRG
EDWARDS D. A., UK
EGOROV Aleksandr I., USSR
EGOROV Yu. V., USSR
EGUCHI K., Japan
EJSMONT Marek, Poland
EL-SABAA Fawzy, Egypt
ELBERT Árpád, Hungary
ELEKES Gy., Hungary
ELIAŠ J., Czechoslovakia
ELSTER Karl-Heinz, GDR
ELSTRODT J., FRG
ELZEIN Fouad, France
EMAMIRAD Hassan, France
ENEDUANYA S. A. N., Nigeria
ENEVA G., Bulgaria
ENGEL Konrad, GDR
ENGEL Wolfgang, GDR

ENGELBERT H. J., GDR
ENGELKING Ryszard, Poland
ENGQUIST B., Sweden
ERDÖS Paul, Hungary
ERIKSSON Folke S., Sweden
ERMAKOV S. M., USSR
ERSHOV A. P., USSR
ERVYNCK G. J. H., Belgium
ESKIN Gregory, USA
EWERT Janina, Poland
EWERT-KRZEMIENIEWSKI Stanisław, Poland

FABIANI P. H., France
FACCHINI A., Italy
FADDEEV L. D., USSR
FÁGEŢAN F. I., Romania
FARID Nabil Zaki, Egypt
FARMANESH Mounesali, Iran
FÄSSLER Albert, Switzerland
FELUCH Barbara, Poland
FENYÖ I., Hungary
FERENS Cezary, Poland
FERNÁNDEZ Cesar L., FRG
FICHTNER K.-H., GDR
FIDELIS Eugeniusz, Poland
FIEDLER Miroslav, Czechosłovakia
FIEDOROWICZ Zbigniew, USA
FIGIEL Tadeusz, Poland
FINAT Javier, Spain
FLATO Moshe, France
FLEISCHMANN Klaus, GDR
FLEMING Wendell H., USA
FLETCHER C. R., UK
FLOREK Jan, Poland
FLORKIEWICZ Bronisław, Poland
FOATA Dominique, France
FOLTYŃSKA Izabela, Poland
FOMENKO A. T., USSR
FONTAINE Jean-Marc, France
FORMELLA Stanisław, Poland
FRAGUELA COLLAR Andres, Cuba
FRANCSICS Gábor, Hungary
FRANK Brigitte, GDR
FRANKIEWICZ Ryszard, Poland
FREDRIKSSON E. H., Netherlands
FREUDENTHAL Hans, Netherlands
FRIEDRICH Thomas, GDR
FRIEDRICH Volkmar, GDR
FRYDRYCHOWICZ Wacław, Poland
FRYSZKOWSKI Andrzej, Poland
FUJITA Akio, Japan
FUKA Jaroslav, Czechoslovakia
FÜREDI Zoltán, Hungary
FURMANOWSKI Tomasz, Poland

GABRIEL P. M. A., Switzerland
GĄDEK Jerzy, Poland
GÄHLER Werner, GDR
GAĬSHUN Ivan V., USSR
GAJDA Adam, Poland
GALLEWICZ Maria, Poland
GAMBARELLI G., Italy
GANCARZEWICZ Jacek, Poland
GARG K. M., Canada
GARGOV G. K., Bulgaria
GÄRTNER Jürgen, GDR
GATIAL Ján, Czechoslovakia
GAUTRIN Henri F., Canada
GAWRON Piotr, Poland
GĘBA Kazimierz, Poland
GĘBAROWSKI Andrzej, Poland
GEHRING, F. W., USA
GEIGLE Werner, FRG*
GENCHEV Todor, Bulgaria
GER Roman, Poland
GERARDIN-LAMÈCHE Khira, France
GERENCSÉR László, Hungary
GERHARDT C., FRG
GEROV Gero, Bulgaria
GERSTENKORN Tadeusz, Poland
GHAFFARI Abdyhassem, Iran*
GHEORGHE G. P., Romania
GICHMAN J. J., USSR
GIEC Tadeusz, Poland
GILBARG D., USA
GINALSKA Stefania, Poland
GINALSKI Czesław, Poland
GIRARD Jean-Yves, France
GIROUX André, Canada
GIRSTMAIR Kurt, Austria
GIUSTI M., France
GŁADYSZ S. S., Poland
GŁAZEK K., Poland
GLEICHGEWICHT B., Poland
GŁOWACKI Paweł, Poland

GLOWINSKI Roland, France
GNATEK Adam, Poland
GNEDENKO B. V., USSR
GODLEWSKI Stanisław, Poland
GOEBEL Kazimierz, Poland
GÖKTÜRK A. Z., Turkey
GOLASIŃSKI Marek, Poland
GOLDSCHMIDT Bernd, GDR
GOLDSTEIN Stanisław, Poland
GOŁDYS B., Poland
GOLEC Janusz, Poland
GOLUBOV Boris, USSR
GOMEZ Francisco, Spain
GONCERZEWICZ Jan, Poland
GONČAR A. A., USSR
GONSHOR Harry, USA
GÖPFERT Alfred, GDR
GORBACHENKO Olga, A., USSR
GORBACHUK M. L., USSR
GORBUNOV V. K., USSR
GORDEZIANI D. G., USSR
GÓRNIEWICZ Lech, Poland
GÓROWSKI Jan, Poland
GÓRSKI Jerzy, Poland
GORZEŃSKI Stanisław, Poland
GOULD Sydney, USA
GRABACKI Jan, Poland
GRABOWSKI Janusz, Poland
GRABOWSKI Marek, Poland
GRAHAM R. L., USA
GRAHAM Victor, W., Ireland*
GRAJEK Czesław, Poland
GRANAS Andrzej, Canada
GRANDE Zbigniew, Poland
GRANT Douglass L., Canada
GRÄSSER H. S. P., South Africa
GRAW Bernd, GDR
GRAY Alfred, USA
GRAY Mary, USA
GREAVES George R. H., UK
GREGUŠ Michal, Czechoslovakia
GREIM Peter, FRG
GRELAK Aleksander, Poland
GREŠAK P., Czechoslovakia
GRIEPENTROG Eberhard, GDR
GRIESS R. L., USA
GRIFFOR E. R., Sweden
GRIGIS Alain, France
GRISHIN Viacheslav N., USSR
GROMADZKI Grzegorz, Poland
GRONAU Detlef, Austria
GROZDEV S., Bulgaria
GRUDA Mieczysław, Poland
GRYTCZUK Aleksander, Poland
GRZĄŚLEWICZ Andrzej, Poland
GRZĄŚLEWICZ Ryszard, Poland
GRZEGORCZYK Piotr, Poland
GRZEGORCZYK Wiesław, Poland
GRZYBOWSKI Jerzy, Poland
GU Chao-Hao, China*
GUADALUPE J. J., Spain
GUDIVOK Peter M., USSR
GUIDY WANDJA Josephine, Ivory Coast
GUIJE W., FRG
GUIRO Abdoulaye, Senegal
GUNNARSSON Thomas E. W., Sweden
GUŚCIORA Henryk, Poland*
GUSTAFSSON A. T., Sweden
GUSTAFSSON Björn S. L., Sweden
GUSYATNIKOV P.B., USSR
GUTLYANSKIĬ Vladimir Ya., USSR
GUTOWSKI Roman, Poland
GYÖNGY István, Hungary
GYÖRI Ervin, Hungary
GYÖRY Kalmán, Hungary

HA HUY Khoai, Viet-Nam
HAARIO H. E. T., Finland
HABA Zbigniew, Poland
HABRAT Stanisław, Poland
HĄCIA L., Poland
HADAYAT Yassaee, Iran
HADZHIEV D. I., Bulgaria
HADZHIIVANOV Nikolai, Bulgaria
HAF Herbert, FRG
HAJDUK Bogusław, Poland
HAJŁASZ Cezary, Poland
HAJNAL András, Hungary
HAJNAL Peter, Hungary
HAJNOSZ Andrzej, Poland
HAJTO Zbigniew, Poland
HALE Jack K., USA
HALE V. W. D., UK
HAMALA Milan, Czechoslovakia
HANSEN J. P., Denmark

HANSEN Sönke, FRG
HART Neal, USA
HARTMAN Stanisław, Poland
HÄRZBECKER Ulrich, FRG
HATORI Asako, Japan
HATTINK J. F., Netherlands
HAVERLIK I., Czechoslovakia
HAYASHI Hiroshi, Japan
HDEIB Hasan, Jordan
HEATH-BROWN D. R., England
HEBERLEIN Krzysztof, Poland
HÉBERT Michel, Canada
HECTOR J. G., France
HEDBERG Lars Inge, Sweden
HEDENMALM P. J. H., Sweden
HEENE Eckehard, GDR
HEGG Ernst,*
HEILIÖ M. I., Finland
HEINRICH S., GDR
HEINZE J., FRG
HEJCMAN Jan, Czechoslovakia
HELLE Renate, GDR
HELMINCK A. G., Netherlands
HELMINCK G. F., Netherlands
HENDERIKS Richard E. D., Netherlands
HENKIN G. M., USSR
HENSZ Ewa, Poland
HERFORT W., Austria
HERMAN Krzysztof, Poland
HERMANN T. Hungary
HERNANDEZ Francisco Luis, Spain
HERNÁNDEZ L. J., Spain
HERRERA ARÁUZ Nelson, Ecuador
HERRIOT John G., USA
HERRIOT Sarah T., USA
HERRMANN Leopold, Czechoslovakia
HERSZBERG J., UK
HESSELINK W. H., Netherlands
HETMAŃSKA Maria, Poland*
HIGUCHI Teiichi, Japan
HINEVA S. T., Bulgaria
HINZ Andreas M., FRG
HIRASHIMA Hideharu, Japan
HITCHIN Nigel J., UK
HOA VU DINH, Viet-Nam
HOANG TUY, Viet-Nam
HÓDI Endre, Hungary
HOFFMANN Frank, GDR
HOHTI Aarno, Finland
HOLASOVA Paula, Czechoslovakia
HOLENDA J. V., Czechoslovakia
HOLEVO A. S., USSR
HOLLY Konstanty, Poland
HOLZAPFEL Rolf-Peter, GDR
HONDA Kin-ya, Japan
HONKASALO H., Finland
HOOLEY Christopher, UK
HÖPPNER R., GDR
HORAK Peter, Czechoslovakia
HORECKI A., USA
HOROZOV E. I., Bulgaria
HOU Zhenting, China*
HOWIE John M., UK
HOWSON A. G., UK
HRISTOV E. H., Bulgaria
HSIANG Wu-Chung, USA
HU THAKYIN, China-Taiwan
HUDZIK Henryk, Poland
HUGGETT S., UK
HULANICKI Andrzej, Poland
HUMKE Paul D., USA
HULEK Klaus, FRG*
HURRELBRINK Jürgen, FRG
HUSKEY H. D., USA
HUSSAIN Mansur, Kuwait
HUYNH MUI, Viet-Nam
HWANG Tea-Yuan, China-Taiwan
HYB W. Z., Poland

IBRAHIM F. N. Fayez, Egypt
ICHIKAWA Kaoru, Japan
IDZIAK Paweł, Poland
IGARÀSHI Masao, Japan
IGNATENKO V. F., USSR
IGODT Paul G. C., Belgium
IITAKA Shigeru, Japan
ILIĆ Ljubica, Yugoslavia
ILIĆ Risto, Yugoslavia
ILIEV Valentin, Bulgaria
IMAI Chuihi, Japan
IMANALIEV Murzalen, USSR
INASSARIDZE H. N., USSR
INGLOT Tadeusz, Poland
IONESCU I. R., Romania
IONITĂ P. C., Romania

IOSIFESCU Marius, Romania
IRTEGOV Valentin D., USSR
ISA BABA Mohammed, Nigeria
ISHIKAWA Susumu, Japan
ISIDRO Jose M., Spain
ISKOVSKIH V. A., USSR
ISMAGILOV Rais, USSR
ITO Yoshihiko, Japan*
IVANOV Aleksandr A., USSR
IVANOV Aleksandr V., USSR
IVANOV N. V., USSR
IVANOV G., Bulgaria
IVANŠIĆ Ivan, Yugoslavia
IVERT P.-A., Sweden
IWAN Andrzej, Poland
IWANAGA Yasuo, Japan
IWANIEC Henryk, Poland
IWANIEC Katarzyna, Poland
IWANIEC Tadeusz, Poland
IWIŃSKI Tadeusz, Poland

JABŁOŃSKI Marian, Poland
JACH Danuta, Poland
JACKOWSKI Stefan, Poland
JACKOWSKI Tomasz, Poland
JACOBSEN E., UN
JACOBSEN Lisa, Norway
JAEGER Arno, FRG
JAFARIAN Ali Akbar, Iran
JAGERS A. A., Netherlands
JAHNKE T., FRG
JAJTE Ryszard, Poland
JAKOBSCHE Włodzimierz, Poland
JAKUBCZYK Bronisław, Poland
JAKUBIAK Mirosław, Poland
JAKUBOWICZ Antoni, Poland
JAKUBOWSKI Adam, Poland
JAKUBOWSKI Tadeusz, Poland
JAKUBOWSKI Zbigniew, Poland
JAMA Danuta, Poland
JAMSHIDIAN Farshid, Iran
JANAS Jan, Poland
JANIAK Teresa, Poland
JANIK Adam, Poland
JANKIEWICZ Maria, Poland
JANKOWSKI Leszek, Poland
JANKOWSKI Tadeusz, Poland
JANKOWSKI Wiktor, Poland
JANSON S., Sweden
JANTZEN J. C., FRG
JANUSZKIEWICZ T., Poland
JARCHOW Hans, Switzerland
JARNICKI Marek, Poland
JARNÍK J., Czechoslovakia
JAROMINEK Władysław, Poland
JAROSZ Janusz, Poland
JAROSZ Krzysztof, Poland
JAROSZEWSKA M., Poland
JARZEMBSKI Grzegorz, Poland
JASKUŁA Janusz, Poland
JAWORSKI Piotr, Poland
JAYNE John E., UK
JĘDRYCZKA Renata, Poland
JĘDRZEJEWSKI Jacek Marek, Poland
JEGIER Jadwiga, Poland
JENDROL Stanislav, Czechoslovakia
JENSEN Ch. V., Denmark
JEŚMANOWICZ L. J., Poland
JEZIERSKI Jerzy, Poland
JIANG Boju, China*
JIMÉNEZ M. A., Cuba
JOHANSSON Bo Ingwar, Sweden
JOHNSON B. E., UK
JOHNSON R. A., USA
JOHNSTONE P. T., UK
JONES P.W., USA
JÓNSSON Jón Hafsteinn, Ireland*
JOSEPH Anthony, Israel*
JÓZEFIAK T., Poland
JUHÁSZ István, Hungary
JUNG Hermann, GDR
JUNGK Werner, GDR
JUREK Z. J., Poland
JURLEWICZ Teresa, Poland
JUST Winfried, Poland

KĄCKI Edward, Poland
KACZOROWSKI Jerzy, Poland
KADEJ Celina, Poland
KAHANE Jean-Pierre, France
KAHANPÄÄ Lauri V., Finland
KAIJSER S. G., Sweden
KAJETANOWICZ Przemysław, Poland
KAJIKAWA Yuji, Japan
KĄKOL Henryk, Poland
KALA R., Poland

KALINA J., Poland
KALINOWSKI Józef, Poland
KALINOWSKI Rafał, Poland
KAMGA Wouafo, Cameroon
KAMIŃSKA Anna, Poland
KAMIŃSKI Andrzej, Poland
KAMIŃSKI Brunon, Poland
KAMONT Zdzisław, Poland
KANIA-BARTOSZYŃSKA Joanna, Poland
KANIE Yukihiro, Japan
KANIEWSKI J., Poland
KANTOR Jean-Michel, France*
KANTOWSKI E. L., USA
KAPCIA Andrzej, Poland
KAPITANSKIĬ Lev V., USSR
KAPOOR O. P., India
KARAFIAT Andrzej, Poland
KARATOPRAKLIEV Georgi, Bulgaria
KARATOPRAKLIEVA-IVANOVA Ivanka, Bulgaria
KAROŃSKI M., Poland
KARP R. M., USA
KARPEEV V. V., USSR
KARWOT Izydor, Poland
KASHIN Boris S., USSR
KASHTANOV V. A., USSR
KASPAROV Guennadi G., USSR
KATONA Gyula, Hungary
KAWAHARA Yusaku, Japan
KAZARYAN G. G., USSR
KAŹMIERCZAK Lech, Poland
KENDEROV P. S., Bulgaria
KETONEN Timo, Finland
KHACHATUROV Vladimir R., USSR
KHAN M. Nazir, Pakistan*
KHATOONABADI M., Iran
KHOVANSKIĬ Askol'd G., USSR
KHUSNUTDINOVA N. V., USSR
KIEŁPIŃSKI Roman, Poland
KIERAT Władysław, Poland
KIIKKA Maire, Finland
KILP Mati A.-B., USSR
KIM Dohan, Korea South
KINOSHITA Michio, Japan
KIRILLOV A. N., USSR
KISIELEWICZ Michał, Poland
KISS Péter, Hungary
KIWIEL K. C., Poland
KLAFSZKY Emil, Hungary
KLAINERMAN Sergiu, USA
KLECHA Tadeusz, Poland
KLEIMAN S. L., Denmark–USA
KLEINER Adam, Poland
KLEINER Witold, Poland
KLEINERT Werner, GDR
KLEMP B. K., Poland
KLIMEK Ewa, Poland
KLIMEK Jacek, Poland
KLIMEK-CHUDY Sławomir, Poland
KLIORYS Constantine, USA
KLIŚ Czesław, Poland
KLOCH Jacenty, Poland
KŁOPOTOWSKI A. A., Poland
KŁOSOWSKA Maria, Poland
KLÖTZLER Rolf, GDR
KLUGE R., GDR
KNAST R., Poland
KNITTER Konrad, Poland
KNOBLOCH H. W., FRG
KOBAYASHI Ugoro, Japan
KOBUS Maria, Poland
KOÇAK Cevdet, Turkey
KOCH H., GDR
KOCHETKOV Dmitrij V., USSR
KOCIŃSKI Jerzy, Poland*
KÖHLER Jürgen, GDR
KOJIMA Sadayoshi, Japan
KOJRO L., Poland
KOŁACZ Henryk, Poland
KOŁAKOWSKI Henryk, Poland
KOLAR Dragan, Yugoslavia
KOLÁŘ Ivan, Czechoslovakia
KOLAROV Dobromir, Bulgaria
KOLAROV Krasimir D., Bulgaria
KOLK Johan A. C., Netherlands*
KOLOTILINA L. Yu., USSR
KOMORNIK Jozef, Czechoslovakia
KONDERAK Jerzy, Poland
KONDRACKI Witold, Poland
KÖNIG Gerhard, FRG
KONOVALOV V. N., USSR
KOPEĆ Ewa, Poland
KOPELL Nancy, USA
KOPIECKI Ryszard, Poland
KOPIŃSKI Jerzy, Poland

KOPOCIŃSKA Ilona, Poland
KOPRINSKY Stephan, Bulgaria
KOPYT Marek, Poland
KOŘÍNKOVÁ Hana, Czechoslovakia
KORNEICHUK N. P., USSR
KORWIN-KOCHANOWSKA Irena, Poland
KOSACHEVSKAYA E. A., USSR
KOŚCIANOWSKI Jan R., Poland
KOSHI Shozo, Japan
KOSIEK Marek, Poland
KOSSAK R., Poland
KOSZUL J. L., France
KÔTA Osamu, Japan
KOTUS Janina, Poland
KOTZE W., South Africa
KOWALCZYK Wojciech, Poland
KOWALEWSKA Zofia, Poland
KOWALSKI Marek, Poland
KOWALSKI Zbigniew S., Poland
KOZAK Michał, Poland
KOZLOV Mikhail V., USSR
KOZŁOWSKA Grażyna, Poland
KOZŁOWSKI Wojciech, Poland
KOŹNIEWSKA Ira, Poland
KRAFT Jürgen, FRG*
KRAJEWSKI Stanisław, Poland
KRAJNAKOVA D., Czechoslovakia*
KRAKOWIAK Wiesław, Poland
KRÁL Josef, Czechoslovakia
KRA'MLI A., Hungary
KRASINKIEWICZ J., Poland
KRASIŃSKA Stefania, Poland
KRASIŃSKI Tadeusz, Poland
KRASNER Marc, France*
KRÄTZEL Ekkehard, GDR
KRAUZE Tadeusz K., Poland
KRECZMAR Antoni, Poland
KRECZNER Robert, Poland
KRÉE Paul, France
KREID Tadeusz, Poland
KRIVENKOV Yu. P., USSR
KRÓLIKOWSKI Andrzej, Poland
KRUPA Andrzej, Poland
KRUPIŃSKI Ryszard, Poland
KRUPSKI Paweł, Poland
KRUSHKAL Samuil L., USSR
KRUSZEWSKA Danuta, Poland
KRYAZHIMSKIĬ A. V., USSR
KRYCH Michał, Poland
KRYGOWSKA Anna Zofia, Poland
KRŽOVSKI Pano, Yugoslavia
KRZYWICKI Andrzej, Poland
KRZYŻ Jan, Poland
KRZYŻEWSKI Karol, Poland
KUBARSKI Jan, Poland
KUBILIUS J. P., USSR
KUBOWICZ Piotr, Poland
KUCHARCZAK Jerzy, Poland
KUCHARCZYK Piotr, Poland
KUCHARSKI Zygfryd, Poland
KUCHARZEWSKI Mieczysław, Poland
KUCIA Anna, Poland
KUĆMIŃSKA Bogusława, Poland
KUCZYŃSKA Ł., Poland
KUDRYAVTSEV L. D., USSR
KUFNER Alois, Czechoslovakia
KUIŃSKI Jacek, Poland
KUIPER Nicolaus, France
KULEJEWSKA E., Poland
KULPA W., Poland
KUMLIN P., Sweden
KÜMMICH H. F., FRG
KUNICKI J. S., Poland
KÜNZI U. M., Switzerland
KUPŚĆ Witold, Poland
KURC W., Poland
KURCIUS Eryk, Poland
KURDYKA Krzysztof, Poland
KURKE Herbert, GDR
KURPIEL Andrzej, Poland
KURYLEV Y. V., USSR
KURŽANSKIĬ A. B., USSR
KURZWEIL Jaroslav, Czechoslovakia
KUTTNER Brian, UK
KUULASMAA Kari, Finland
KUZAK Włodzimierz, Poland
KUZNETSOV N. V., USSR
KUZNETSOV S. E., USSR
KUZNETSOV Yu. A., Gorkiĭ, USSR
KUZNETSOV Yu. A., Moskva, USSR
KVEDARAS B. V., USSR
KWAPISZ Marian, Poland
KWIATKOWSKI Jan, Poland

LABUDA Iwo M., Poland

LACEY H. E., USA
LACKSCHEWITZ C. B., West Berlin
ŁADA Andrzej, Poland
LADYŽENSKAYA O. A., USSR
LAI Hang-Chin, China-Taiwan
LANGOV Anani, Bulgaria
LAPIŇŠ Janis, USSR
LARNER D. J., Netherlands
LASOTA Andrzej, Poland
LASSAK Marek, Poland
LASSNER Gerd, GDR
LAU (Anthony) To-Ming, Canada
LAUDAL O. A., Norway
LAVRENTIEV Mikhail, USSR
ŁAWRYNOWICZ Julian, Poland
LAX P. D., USA
LAZAROVA B., Bulgaria
ŁAZARSKA Katarzyna, Poland
LAZOV R. G., Bulgaria
LEBEDEV V. I., USSR
LEBORGNE Daniel, France*
LE DUNG TRANG, Viet-Nam
LÊ VAN THIÊM, Viet-Nam*
LECUYER M. J., France
LEDWINA Teresa, Poland
LEE Peter Martin, UK
LEHTINEN Matti, Finland
LEHTINEN Markku S., Finland
LEHTO Olli, Finland
LEHTO Pertti, Finland
LEINDLER László, Hungary
LEITERER Jürgen, GDR
LEMAŃCZYK Mariusz, Poland
LEMPERT László, Hungary
LENGYEL Tamás, Hungary
ŁEŃSKI Włodzimierz, Poland
LENZING Helmut, FRG
LEONTIEV V. K., USSR
LEROY André, Belgium
LESZCZYŃSKA Zofia, Poland
LESZCZYŃSKI Bogumił, Poland
LETICHEVSKY A. A., USSR
LEVI R. V., Bulgaria
LEWANDOWSKI G., Poland
LEWANOWICZ S., Poland
LEWIS Wayne, USA
LEWKOWICZ Marek, Poland
LIBERA Richard, USA
LIESE Friedrich, GDR
LIN Wen Hsiung, China-Taiwan
LINDE Werner, GDR
LINDSTRÖM I., Sweden
LINTON F. E. J., USA
LIONS J.-L., France
LIONS P.-L., France
LIPCSEY Zsolt, Hungary
LIPECKI Zbigniew, Poland
LIPIŃSKI J. S., Poland
LIPPOLD G., GDR
LIPSKI Tadeusz, Poland
LISIECKI Wojciech, Poland
LISSOWSKI Andrzej, Poland
LITEWSKA Katarzyna, Poland
LITVINCHUK G. S., USSR
LIU Shih-Chao, China-Taiwan
LIU Yingming, China*
LIZAK Romualda, Poland
LJUBENOVA-TODOROVA Elena, Bulgaria
LLUIS Emilio, Mexico
LOEB Peter A., USA
ŁOJASIEWICZ S., Poland
ŁOJASIEWICZ S. O., Poland
ŁOMNICKI Adam, Poland
LÓPEZ Guillermo, Cuba
LÓPEZ-MORALEZ Martin, Cuba
LORCH Lee, Canada
LORENTZ Krzysztof, Poland
LOSHITSKIĬ A. F., USSR
LOUHIVAARA Ilppo Simo, Finland
LOUTON T. K., USA
LOVÁSZ D. László, Hungary
LOVÁSZ László, Hungary
LOVE Eric Russel, Australia
LOVICAR V., Czechoslovakia
LOZANOV Čavdar, Bulgaria
LUAN Ru-Shu, China*
LUBAŃSKI Mieczysław, Poland
ŁUCZAK Andrzej, Poland
ŁUCZAK-KUMOREK Elżbieta, Poland
ŁUKASZEWICZ Grzegorz, Poland
ŁUKASZEWICZ Józef, Poland
LUKEŠ Jaroslav, Czechoslovakia
LUKSHIN A. V., USSR
LUSZTIG G., USA
LUUKKAINEN Jouni, Finland

L'VOVA M. E., USSR
LYKOVA O. B., USSR
ŁYSIK Grzegorz, Poland

MACEDOŃSKA-NOSALSKA Olga, Poland
MACINTYRE Angus J., UK
MACK John, USA
MAĆKOWIAK Tadeusz, Poland
MACPHERSON Robert D., USA
MĄCZKA Czesław, Poland
MĄCZYŃSKI Maciej, Poland
MADAJCZYK J. L., Poland
MADGUEROVA Andreana S., Bulgaria
MĄDRECKI Andrzej, Poland
MAGDOŃ Maria, Poland
MAGIERA Józef, Poland
MAGNÚSSON J., Iceland
MAI Brigitte, GDR
MAJCHER Genowefa, Poland
MAJCHROWSKI Marian, Poland
MAJEWSKI Wł. A., Poland
MAJUMDAR Subrata, Bangladesh
MAKAGON Andrzej, Poland
MAKAROV Vitaly, USSR
MAKARY Farouk, Libya
MAKHNIOV Aleksander, USSR
MAKOWIECKA H., Poland
MĄKOWSKI Andrzej, Poland
MALAWSKA A., Poland
MALAWSKI M., Poland
MALCZAK Jan, Poland
MALEC Marian, Poland
MALIGRANDA Lech, Poland
MALINOVSKIĬ V. K., USSR
MALLIAVIN Paul, France
MALTSINIOTIS G., France
MAMOURIAN Ahmad, Iran
MAMRILLA Juraj, Czechoslovakia
MAN Chi-Sing, Canada
MANDELBROT Benoit B., USA
MANDZHAVIDZE G. F., USSR
MANDL Petr, Czechoslovakia
MAÑÉ Ricardo, Brazil
MANIAKOWSKI Feliks, Poland
MANIN Yu. I., USSR
MANKIEWICZ Piotr, Poland
MANOLOV Spas, Bulgaria
MARCELLAN Francisco, Spain
MARCHENKOV S. S., USSR
MARCHIONNA E., Italy
MARCHIONNA TIBILETTI C., Italy
MARCHOW Ewa, Poland
MARCHUK N. G., USSR
MARCINIAK Andrzej, Poland
MARCINKOWSKA Hanna, Poland
MARINESCU G., Romania
MARINOV Marin, Bulgaria
MÁRKI László, Hungary
MARKOV Svetoslav, Bulgaria
MARKOV V. V., USSR
MARKS Tomasz, Poland
MARLEWSKI Adam, Poland
MARTENS Henrik H., USA*
MARTINEZ-MARTINEZ Antonio, Spain
MARTINEZ Servet, Chile
MARTYNYUK A. A. USSR
MARUŠIAK Pavol, Czechoslovakia
MARUYAMA Akira, Japan
MÄRZ Roswitha, GDR
MARZANTOWICZ Wacław, Poland
MASAMICHI Tomida, Japan
MASANI Pesi R., USA
MAŚLANKA Paweł, Poland
MASLENNIKOVA V. N., USSR
MASLOV Viktor P., USSR
MASON R. C., UK
MASSER David W., UK
MASTALERZ-WAWRZYŃCZAK Maria, Poland
MASTYŁO N., Poland
MASUDA Mikiya, Japan
MASUDA Tetsuya, Japan
MATÉ László, Hungary
MATHIA Zdzisław, Poland
MATHIEU Philippe, Belgium*
MATKOWSKI Janusz, Poland
MATŁOKA Marian, Poland
MATOV Vassil M., Bulgaria
MATSUMOTO Koji, Japan*
MATTHES Klaus, GDR
MATTILA Kirsti M., Sweden
MATTILA Pertti E. J., Finland
MATUSKI Zdzisław, Poland
MATUSZEWSKA Wanda, Poland
MATVEEV V. F., USSR

MAZUR Barry, USA
MAZUR Tomasz, Poland
MAZUROV Viktor, USSR
MAZZIOTTO G., France
MCBETH R., UK
MCCRIMMON K. M., USA
MCDONOUGH T. P., UK
MCINTOSH Alan G. R., Australia
MEKHMEDOV H., Bulgaria
MĘCZARSKI Marek, Poland
MEDVEDEV N. Ya., USSR
MEGYESI László, Hungary
MEINEL Christoph, GDR
MEISTERS G. H., USA
MEJZA Stanisław, Poland
MERZBACH Ely, Israel
MEYER Yves, France
MICCHELLI Charles, USA
MICHAEL Jerzy Henryk, Poland
MICHALAK Danuta, Poland
MICHALAK Jacek, Poland
MICHEL H., GDR
MICHNIEWSKI Jerzy, Poland
MICHNIKOWSKA-PLUCIŃSKA Teresa, Poland
MICZKO Antoni, Poland
MIDURA St., Poland
MIERCZYŃSKI Janusz, Poland
MIGOŃ J., Poland
MIHALACHE Nicolae, Romania
MIHALJINEC Mirko, Yugoslavia
MIKLÓS Dezsö, Hungary
MIKLOS Stanisław, Poland
MIKLOŠKO Jozef, Czechoslovakia
MIKOŁAJCZAK J., Poland
MIKOŁAJCZYK Leon, Poland
MIKOLÁS Miklós, Hungary
MIKOS Henryk, Poland
MIKOSZ Mirosława, Poland
MIKULSKI Piotr, USA
MIKUSIŃSKI Jan, Poland
MIKUSIŃSKI P., Poland
MILEWSKI Jan, Poland
MILKA A. D., USSR
MILLER D. D., USA
MILLETT K. C., USA
MILLIONSHCHIKOV V. M., USSR
MILNES Paul, Canada
MIŁO Władysław, Poland
MILOSLAVSKY G. C., USA
MINCER Bohdan, Poland
MINGARELLI Angelo, B., Canada
MIODUSZEWSKI Jerzy, Poland
MIRAZHCHIJSKI I. S., Bulgaria
MISHCHENKO A. S., USSR
MISIEWICZ Jolanta K., Poland
MISIUREWICZ Michał, Poland
MITROPOLSKIĬ Yu. A., USSR
MITSUMATSU Yoshihiko, Japan
MIZUKAMI Masumi, Japan
MIZUTANI Tadayoshi, Japan
MODROW H. D., GDR
MOGHADDAM M. R. R., Iran
MOHRHERR J., USA
MOHRY Benno, FRG
MOKOBODZKI G., France
MOLLIN Richard A., Canada
MOLNAR Sandor, Hungary
MOLZAN Burkhard, GDR
MORALES Oscar, Peru
MORALES-LUNA Guillermo, Mexico
MORCHAŁO Jarosław, Poland
MORI Shigefumi, Japan
MOROZAN T., Romania
MOROZOV Nikita F., USSR
MOROZOV Vladimir P., USSR
MOSER Jürgen, Switzerland
MOSTOW George Daniel, USA
MOSTOWSKI A. W., Poland
MOSTOWSKI T., Poland
MOSZNER Zenon, Poland
MOTORNYĬ V. P., USSR
MOZGAWA Witold, Poland
MROZEK M. T., Poland
MÜLLER P. F. X., Austria
MÜLLER Werner, GDR
MURAWSKI R., Poland
MURAYAMA M., Japan
MURTY U. S. R., Canada
MUSHKAROV Oleg, Bulgaria
MUSIAŁ K., Poland
MUSIELAK Anna, Poland
MUSIELAK Helena, Poland
MUSIELAK Julian, Poland
MUSZYŃSKI Jerzy, Poland
MUTAFCHIEV L. R., Bulgaria

MYJAK Józef, Poland
MYSIOR Adam, Poland
MYSZEWSKI Jan, Poland

NÄÄTÄNEN M., Finland
NAGAEV A. V., USSR
NAGY Zs., Hungary
NAITO Tadao, Japan
NALBACH Małgorzata, Poland
NANDA V. C., India
NARASIMHAN M. S., India
NARCOVICH Francis J., USA
NARKIEWICZ Władysław, Poland
NASSOPOULOS George F., Greece
NATKANIEC Tomasz, Poland
NATSUME Toshikazu, Japan
NAUMANN Joachin, GDR
NAZAROVA L. A., USSR
NEDEVA Cvetanka, Bulgaria
NĘDZA-KUBINIEC Halina, Poland
NEHRING Jerzy, Poland
NEIDHARDT Hagen, GDR
NĚMEC Petr, Czechoslovakia
NEMOGA Karol, Czechoslovakia
NEŠETŘIL Jaroslav, Czechoslovakia
NETUKA Ivan, Czechoslovakia
NEUBAUER G. J., FRG
NEUBRUNN Tibor, Czechoslovakia
NEUMAN Edward, Poland
NEUMANN Bernhard H., Australia
NEUMANN Olaf, GDR
NEUMANN Peter, GDR
NEWMAN Michael F., Australia
NGO VIET TRUNG, Viet-Nam
NGUYEN CANG, Viet-Nam
NGUYEN CUONG, Poland
NGUYEN TO NHU, Viet-Nam
NGUYEN XUAN TAN, Viet-Nam
NGUYEN ZUI TIEN, Viet-Nam
NICOLAS Anne-Marie, France
NICOLAS Jean-Louis, France
NICOLOV Peyo, Bulgaria
NICULESCU C., Romania
NICHEVA Maria M., Bulgaria
NIECHWIEJ Jan, Poland
NIEDERHAUSEN H., Canada
NIEDOBA Wiesław, Poland
NIEMI H. O., Finland
NIEMYJSKA Wiesława, Poland
NIETO J. I., Canada
NIEUWAŻNY Wacław Paweł, Poland
NIEZGÓDKA Marek, Poland
NIKIEL Jacek, Poland
NIKODEM Kazimierz, Poland
NIKOLAEV Vladimir, S., USSR
NIKOLOVA D. B., Bulgaria
NIKOL'SKIĬ M. S., USSR
NIKOL'SKIĬ S. M., USSR
NILSSON P. G., Sweden
NINNEMANN O., FRG
NISHIDA T., Japan*
NIWIŃSKI Damian, Poland
NOACK Monika, GDR
NORWA Jerzy, Poland*
NOURI-MOGHADAM M. R., Iran
NOVÁK Bretislav, Czechoslovakia
NOVÁKOVÁ Eva, Czechoslovakia
NOVOKSHENOV V. Yu., USSR
NOWAK Andrzej, Katowice, Poland
NOWAK Andrzej, Wrocław, Poland
NOWAK Barbara, Poland
NOWAK Bogdan, Poland
NOWAK Slawomir, Poland
NOWAKOWSKI Andrzej, Poland
NOWECKI Bogdan, Poland*
NOWICKA Krystyna, Poland
NOWICKI Andrzej, Poland
NOWIŃSKI Krzysztof S., Poland
NOZAKI Yasuo, Japan
NUMAKURA Katsumi, Japan

OBRETENOV Apostol, Bulgaria
ODOŻYŃSKI Ryszard, Poland
ODZIJEWICZ Anatol, Poland
OESTERLÉ, Joseph, France
OGUS A., USA
OHTA Minolu, Japan
OKADA Masami, Japan
OKB EL-BAB A. S., Egypt
OKNIŃSKI Andrzej, Poland
OKNIŃSKI Jan, Poland
OKRASIŃSKI Wojciech, Poland
ØKSENDAL B., Norway
OKTABA W., Poland
OLECH Czesław, Poland
OLĘDZKI Wiesław, Poland

OLEĬNIK Olga, USSR
OLEJNICZAK Andrzej, Poland
OL'ŠANSKIĬ A. Y., USSR
OLSZAK Zbigniew, Poland
OMAR A. A., Egypt
OMBACH Jerzy, Poland
OMILJANOWSKI Krzysztof, Poland
OMORI Michio, Japan
OPOROWSKA Joanna, Poland
OPOROWSKI Bogdan, Poland
OPOZDA Barbara, Poland
OREVKOV V. P., USSR
ORLICZ W., Poland
ORŁOWSKI Norbert, Poland
OSADA Józef, Poland
OSHIMA Toshio, Japan
OSKOLKOV A. P., USSR
OSŁOWSKI Bogdan, Poland
OSTASZEWSKI Adam, UK
OSTERBURG J., USA
OSTIANU Natalia M., USSR
OSWALD Peter, GDR
OTELBAEV Moukhtarbai, USSR
OTTENBERG Rudolf, GDR
OVERSTEEGEN Lex G., USA
ÖZLÜK Ali E., Turkey

PACH J. G., Hungary
PACHOLSKI Leszek, Poland
PÄIVÄRINTA Lassi Juhani, Finland
PAŁASIŃSKI Marek, Poland
PALCZEWSKI Bolesław, Poland
PÁLFY Peter Pal, Hungary
PALIS Jacob Jr, Brazil
PAŁKA Janina, Poland
PALKA Zbigniew, Poland
PANAHI Tabrizi F., Iran
PANCHEVA E., Bulgaria
PANKIEWICZ Stanisława, Poland
PANKOV A. A., USSR
PAPADIMA S., Romania
PAPIER E., France
PAPUC Dan, Romania
PARCZYK Krystyna, Poland
PARDAŁA Antoni, Poland
PARSHIN A. N., USSR
PARTHASARATHY Rajagopalan, India
PASCU E., Romania
PASICKI Lech, Poland
PASNICU C., Romania
PASTERNAK-WINIARSKI Zbigniew, Poland
PASZEK Gerard, Poland
PASZKIEWICZ Adam, Poland
PÁSZTOR Jozsef, Hungary
PATKOWSKA Hanna, Poland
PATTERSON E. M., UK
PAUNESCU L., Romania
PAVLOV B. S., USSR
PAVLOV Yu. L., USSR
PAWLIKOWSKI Janusz, Poland
PAWŁUCKI Wiesław, Poland
PAZDERSKI Gerhard, GDR
PAYNE Treva N., USA*
PEART P. B., Jamaica
PELCZAR Andrzej, Poland
PEŁCZYŃSKI Aleksander, Poland
PELLA Adam, Poland
PELLAUMAIL J., France
PELLETIER Jooan M. W., Canada
PENKOV Ivan B., Bulgaria
PERADZYŃSKI Zbigniew, Poland
PEREŃSKI Sławomir, Poland
PETHÖ Attila, Hungary
PETKOV Milko, Bulgaria
PETKOV Vesselin, Bulgaria
PETKOVA S. V., Bulgaria
PETREHUS I. V., Romania
PETROSYAN L. A., USSR
PETROV P., Bulgaria
PETRUSHEV P. P., Bulgaria
PETZ Dénes, Hungary
PEYRIÈRE J., France
PFISTER Gerhard, GDR
PHILIPPIN G. A., Canada
PHILIPPOU Andreas N., Cyprus
PHILLIPS Ralph S., USA
PIASECKI Krzysztof, Poland
PICCARD Sophie, Switzerland
PIECHOCKI Władysław, Poland
PIELICHOWSKI Wacław, Poland
PIENE Ragni, Norway
PIEPERS Jan, Netherlands
PIERZCHALSKI Antoni, Poland
PIESYK Zenon, Poland
PIETRZYK Barbara, Poland

PIETSCH A., GDR
PIHKO Jukka, Finland
PIŁAT Barbara, Poland
PILLAT U., GDR
PIMSNER M. V., Romania
PINČUK S. I., USSR
PINO-NIÑO Alvaro, Panama
PINTZ J., Hungary
PIÓREK Józef, Poland
PIÓRO M. P., Poland
PIOTROWSKI Walerian, Poland
PIRINEN Aulis, Finland
PISIER Gilles, France
PISKOREK Adam, Poland
PLÁ Héctor, Cuba
PLAKSIN Vladimir A., USSR
PŁATKOWSKI Janusz, Poland*
PLATONOV Vladimir P., USSR
PLATT Czesław, Poland
PLAUMANN Peter, FRG
PLAZA Jan, Poland
PLEASANTS Peter A. B., UK
PLEŚNIAK Wiesław, Poland
PLESNÍK Ján, Czechoslovakia
PLEWA Mieczysław, Poland
PLIŚ Andrzej, Poland
PLISS V. A., USSR
PŁONKA Ernest Erwin, Poland
PŁOSKI Arkadiusz, Poland
PLUCIŃSKA Agnieszka, Poland
PLUCIŃSKI Edmund, Poland
POGODA Zdzisław, Poland
POGORELOV A. V., USSR
POGORZAŁY Zygmunt, Poland
POGORZELSKI H. A., USA
POGOSYAN Suren, USSR
POKRZYWA Andrzej, Poland
POL Elżbieta, Poland
POL Roman, Poland
POLICHT Jerzy, Poland
POLKOWSKI L. T., Poland
POLLINGTON Andrew D., UK
POLOVINKIN E. S., USSR
POMMARET J.-F., France
POORNIMA S., India
POPA S. T., Romania
POPENDA Jerzy, Poland
POPIVANOV Nedju, Bulgaria
POPIVANOV Peter, Bulgaria
POPOV Angel, Bulgaria
POPOV Georgi, Bulgaria
POPOV Mikhail M., USSR
POPOV Vasil, Bulgaria
POPOWICZ Sławomir, Poland
PORADA M., Poland
POREDA W. J., Poland
PORTENKO N. I., USSR
PORTEOUS Hugh L., UK
PORUBSKÝ Štefan, Czechoslovakia
POSTAWA Stanisława, Poland
POTAPOV M. K., USSR
POWĄZKA Zbigniew, Poland
POWELL M. J. D., UK*
PRAGACZ Piotr, Poland
PRESMAN E. L., USSR
PRICE M. C., Netherlands
PRIVALOV A. A., USSR
PRODANOV Ivan, Bulgaria
PROKHOROV Yu. V., USSR
PROSKURIN N. V., USSR
PROSKUROWSKI W., USA
PRÖSSDORF Siegfried, GDR
PRÓSZYŃSKI Andrzej, Poland
PRUCNAL Tadeusz, Poland
PRUDNIKOV A. P., USSR
PRUSZKO Tadeusz, Poland*
PRZYBYCIN J., Poland
PRZYŁUSKI K. M., Poland
PRZYTYCKI Feliks, Poland
PRZYTYCKI J. H., Poland
PTAK Marek, Poland
PTÁK Pavel, Czechoslovakia
PUCCI Carlo, Italy
PULTAR Milan, Czechoslovakia
PURISCH Steven, USA
PURKERT Walter, GDR
PUTINAR M., Romania
PYTLIK Tadeusz, Poland

QI Min-you, China*
QUASTHOFF U., GDR
QUEK TONG SENG, Singapore

RABCZUK Rościsław, Poland
RABIJEWSKA Barbara, Poland
RABIN Micheal, Israel*

RACHEV Svetlozar, Bulgaria
RĄCZKA Ryszard, Poland
RADENSKY Atanas, Bulgaria
RADEV S., Bulgaria
RADEVA R. K., Bulgaria
RADOSLOVESCU Anca, Romania
RADU Nicolae, Romania
RADZHABOV N. R., USSR
RADZISZEWSKI Konstanty, Poland
RAJAONA Pierre Rius, Madagascar
RAJBA Teresa, Poland
RAKHMANOV E. A., USSR
RALSTON Anthony, USA
RAMADANOV Ivan Pierre, Bulgaria
RAMAZANOV M. D., USSR
RANGELOV Cviatko, Bulgaria
RAPHAEL L. A., USA
RAPINCHUK A. S., USSR
RASOULIAN Amid, Iran
RASSIAS George, Greece*
RASSIAS Themistocles, Greece
RATAJSKI Henryk, Poland
RATSA M. F., USSR
RAUTENBERG Wolfgang, FRG
RĂUŢU Gheorghe, Romania
RAY Ajit Kumar, Canada*
RAY N., UK
RAZZAGHI-KASHANI Mir Mehdi, Iran
RAZZAGHI-KASHANI Mohsen, Iran
READE Maxwell O., USA
RECILLAS-JUAREZ Félix, Mexico
RECZEK Krzysztof, Poland
REES C. S., USA
REES S. M., UK
REGIŃSKA Teresa, Poland
REIS Raquel M., Portugal
REITEN Idun, Norway
REIZINŠ Linards E., USSR
REMESLENNIKOV V. N., USSR
REMPAŁA R., Poland
RESHETNYAK Yu. G., USSR
RESPONDEK Witold, Poland
REUTER Dietrich, FRG
RIAZA R., Spain
RIBET K. A., USA
RICHMOND Bruce, Canada*
RIEČAN Beloslav, Czechoslovakia
RIEČANOVÁ Zdena, Czechoslovakia
RIESEL H. I., Sweden
RIIHENTAUS Juhani, Finland
von RIMSCHA Michael, FRG
RINGEL C. M., FRG
RISNES Martin, Norway
RODRIGUEZ-SALINAS Baltasar, Spain
ROE J., UK
ROESLER F., FRG
ROGAI Eliferie, Romania
ROGULSKI J. S., Poland
ROJAS Rául O., FRG
ROJEK Tadeusz, Poland
ROLSKI Tomasz, Poland
ROMPAŁA Władysław, Poland
RÓNYAI Lajos, Hungary
ROOS Guy, France
ROSENBAUM Kurt, GDR
ROSSOWSKA Michalina, Poland
ROSU R. M., Romania
ROTER Witold, Poland
ROTHBERGER Fritz, Canada*
ROTHE Detlef, GDR
ROTKIEWICZ Andrzej, Poland
ROUXEL Bernard, France*
ROWECKA Ewa, Poland
ROXIN Emilio O., USA
ROZANOV Yu. A., USSR
RÓZSA Pál, Hungary
RUBAN V. I., USSR
RUCIŃSKI A., Poland
RUDAK Leszek, Poland
RUDNIK Krzysztof, Poland
RUELLE D., Belgium/France
RUMAK Tadeusz, Poland
RUMYANTSEV B. N., USSR
RUSEK Barbara, Poland
RUSEK Kamil, Poland
RUSINEK Jan, Poland
RUSZCZYK Henryk, Poland
RUSZCZYK Irena, Poland
RUSZCZYŃSKI Andrzej, Poland
RUTKOWSKI Andrzej K., Poland
RUTTER J. W., UK
RUZSA Imre Z., Hungary
RYBNIKOV A. K., USSR
RYCERZ Anna, Poland
RYCHLEWSKI Jan, Poland
RYCZAJ J., Poland

RYDZEWSKI Piotr, Poland
RYLL-NARDZEWSKI Czesław, Poland*
RYSHKOV S. S., USSR
RYTERSKI Jurand, Poland
RYTTER W., Poland
RYZNAR Michał, Poland

SABBAH Claude, France
SACHS H., GDR
SADOWSKI Michał, Poland
SADOWSKI Paweł, Poland
SAKAGAMI Eisho, Japan
SAKANE Hizuru, Japan
SAKOTO Moussa, Ivory Coast
SALAKHITDINOV M., USSR
ŠALÁT Tibor, Czechoslovakia
SALAVOVÁ Katarina, Czechoslovakia
SALBANY S., Zimbabwe
SALBERGER P. E., Sweden
SALINAS L., Chile
SALLAM Samir, Egypt
SAMOĬLENKO V. G., USSR
SAMOTIJ K., Poland
SANDS A. D., UK
SANIKIDZE D. G., USSR
SANJUÁN GONZÁLEZ Julio Cesár, Poland
SANJURJO J. M. R., Spain
SAPAGOVAS M. P., USSR
SÄRKIJÄRVI J. J., Finland
SÁRKÖZY A., Hungary
SATIMOV Nouman Yu., USSR
SATO H., Japan
SATO Mikio, Japan
SATO Yasuko, Japan
SATORU Watari, Japan
SAWOŃ Zbigniew, Poland
SAZHENKOV A. N., USSR
SCHEERER Hans, FRG*
vom SCHEIDT Jürgen, GDR
SCHENZEL Peter, GDR
SCHIFFELS Gerhard, FRG
SCHILL Walter, FRG
SCHINZEL Andrzej, Poland
SCHIPP Ferenc, Hungary
SCHIRMER Helga, Canada
SCHMICKLER-HIRZEBRUCH Ulrike, FRG
SCHMIDT A. L., Denmark
SCHMÜDGEN Konrad, GDR
SCHNEIDER Klaus, GDR
SCHNEIDER Manfred, GDR
SCHNEIDER Michael, FRG
SCHOCH Manfred, GDR
SCHUDER Werner, FRG
SCHULZE Bert-Wolfgang, GDR
SCHÜLZE Volker, FRG
SCHUPPLI Reto, Switzerland
SCHWABIK Štefan, Czechoslovakia
SCHWARTZ Jacob T., USA
SCHWEIGERT Dietmar, FRG*
SCHWEITZER Paul A., Brazil
ŠEDA Valter, Czechoslovakia
SEDLÁČEK Jiři, Czechoslovakia
SEDRAK Khairy Amiy, Egypt
SĘDZIWY Stanisław, Poland
SEESE D. G., GDR
SEGSCHNEIDER J., FRG
SEIFERT B. G., France
SEKINE Tadayuki, Japan
SELENIUS Clas-Olof, Sweden
SELL George R., USA
SEMADENI Zbigniew, Poland
SENDLEWSKI Andrzej, Poland
SENDOV Blagovest, Bulgaria
SENDOV Bozhidar, Bulgaria
SERAFIN Stanisław, Poland
SEROV V. S., USSR
SERRIN James B., USA
SEVENSTER A., Netherlands
SHABUNIN M. J., USSR
SHADMAN D., Iran
SHAHSMAHANI Mirshams, Iran
SHAHVARANI Ahmad, Iran
SHAĬDUROV V. V., USSR
SHAKHBAZYAN K. V., USSR
SHANIN N. A., USSR
SHARKO Vladimir, USSR
SHARKOVSKIĬ A. N., USSR
SHARMA Banwarilal, India
SHAW Sen-Yen, China-Taiwan
SHELAH Saharon, Israel*
SHEMETKOV L. A., USSR
SHESTAKOV Ivan P., USSR
SHEVCHUK I. A., USSR
SHIBATA Katsuyuki, Japan

SHIELDS A., USA
SHINOHARA Masahiko, Japan
SHIRAKI Mitsunobu, Japan
SHIRINBEKOV Mirzosan, USSR
SHIZUMA Ryoji, Japan
SHOPOLOV N. N., Bulgaria
SHORE R. A., USA
SHTOGRIN M. I., USSR
SIAFARIKAS Panagiotis, Greece
SIBIRSKIĬ K. S., USSR
SICIAK Józef, Poland
SIDDIQI Jamil A., Canada
SIDOROV A. F., USSR
SIDZ Leszek, Poland
SIEBER Norbert, GDR
SIEJKA Henryka, Poland
SIEKLUCKI Karol, Poland
SIERPIŃSKA Anna, Poland
SIKORSKI A. K., Poland
SIKORSKI Maciej W., Sweden
SILIN D. B., USSR
SIMÁNYI Nándor, Hungary
SIMON Leon, Australia
SIMONOVITS Miklós, Hungary
SIMSON Daniel, Poland
SINGMASTER David USA/UK
SINGH U. N., India
SINHA S. R., India
SINYUKOV N. S., USSR
SIPE P. L., USA
ŠIRÁŇ Jozef, Czechoslowakia
SIU Yum-Tong, USA
SIVERTSEN P., Norway
SIWEK Edward, Poland
SIWEK Helena, Poland
SJÖGREN P., Sweden
SKALMIERSKI Bogdan, Poland
SKALSKA Krystyna, Poland
SKIBIŃSKI Przemysław, Poland
SKORDEV Gencho, Bulgaria
SKÓRNIK Krystyna, Poland
SKOWROŃSKA Mirosława, Poland
SKOWROŃSKI Andrzej, Poland
SKRYPNIK I. V., USSR
SKRZYPCZAK Leszek, Poland
SKUBACHEVSKIĬ A. L., USSR
SKULIMOWSKI A. M. J., Poland
SKUPIEŃ Zdzisław, Poland
SKURIKHIN E. E., USSR
SKWIERZ Aleksander, Poland
SŁABY Marek, Poland
SŁADEK Andrzej, Poland
SŁAWIK Lech, Poland
SLISENKO A. O., USSR
ŚLIWA Tadeusz, Poland
SŁOMCZYŃSKI Wojciech, Poland
SŁOMIŃSKA Jolanta, Poland
SŁOMIŃSKI Józef, Poland
SŁOMIŃSKI Leszek, Poland
SŁOMSKA Albina, Poland
SMAILOV E. S., USSR
SMAJDOR Andrzej, Poland
SMAJDOR Wilhelmina, Poland
SMALØ Sverre O., Norway
ŚMIAŁEK Henryka, Poland
SMÍTAL Jaroslav, Czechoslovakia
SMITH A. G., UK
SMITH John H., USA
SMITH Frank P., USA*
SMOLEŃSKI W. H., Poland
SMOLUK Antoni, Poland
SNEDDON Ian N., UK
SNOHA Lubomír, Czechoslovakia
SOBOL R., Poland
SOBOLEV Vladimir A., USSR
SOBOLEWSKI Mirosław, Poland
SOCHA Lesław, Poland
SOCHILINA A. V., USSR
SODOWYN Batsuren, Mongolia
SOFONEA M., Romania
SOKOŁOWSKI Jan, Poland
SOLAK Wiesław, Poland
SOLARIN A. R. T., Nigeria
SOLEV Valentin N., USSR
SOLOVEV Aleksandr D., USSR
SOŁTYSIAK Andrzej, Poland
SOROKIN Anatolij D., USSR
SÓS V., Hungary
SOSNOWSKI J. S., Poland
SOSNOWSKI Z., Poland
SOSULSKI Władysław, Poland
SOUČEK Jiři, Czechoslovakia
SOULÉ Christophe, France
SPAKOWSKI Andrzej, Poland
SPĂTARU Aurel, Romania
SPENCER J., USA

SPIEŻ S., Poland
SRZEDNICKI Roman, Poland
STACHURSKA Barbara, Poland
STAGNARO Ezio, Italy
STANEK Stanisław, Poland
STANEVICHIUS A. J., USSR
STANGLER Sieghart, FRG
STANILOV Grosio, Bulgaria
STANISZEWSKA Janina, Poland
STANKIEWICZ Jan, Poland
STANLEY R., USA
STAPP L., Poland
STASICA Jacek, Poland
STATULEVIČIUS V. A., USSR
STAVROVA D. N., Bulgaria
STEINER H. G., FRG
STEINER Richard, UK
STEMPAK Krzysztof, Poland
STEPANOV S. A., USSR
STEPANOV V. D., USSR
STĘPNIEWSKI Alfred, Poland
STERNHEIMER Daniel, France
STETTNER Łukasz, Poland
STEVENS J., Netherlands
STOBBE Helmut, GDR
STOCHEL Jan, Poland
STOCHEL Jerzy, Poland
STOER J., FRG
STÖHR Ralph, GDR
STOILOV Peyo, Bulgaria
STOJKOVIĆ Neda, Yugoslavia
STOLARCZYK Maria, Poland
STOLTENBERG-HASEN Viggo, Sweden
STOPA Jerzy, Poland
STOYANOV J., Bulgaria
STOYANOV L. N., Bulgaria
STRABURZYŃSKI Jan, Poland
STRAUME E. J., Norway
STRAUSS D., UK
STREULI E., Switzerland
STRÖMBERG J. O., Sweden
STUART Allan E. G., UK
STÜCKRAD J., GDR
STUDNIARSKI Marcin, Poland
STUDZIŃSKI Krzysztof, Poland
STYGAR Danuta, Poland
STYŚ Tadeusz, Poland
SUCHANECKI Zdzisław, Poland
SUCHENEK Marek, Poland
SUDOLSKI Jan, Poland
SUDŽIUTE Daina P., USSR
SULANKE R., GDR
SULIŃSKI Adam, Poland
SULTANAEV Yaudat T., USSR
SULTANGAZIN Umirzak M., USSR
SUNG C. H., China-Taiwan
SURMANIDZE O. E., USSR
SUZUKI Komei, Japan
ŠVARC Rudolf, Czechoslovakia
ŠVEC Marko, Czechoslovakia
SVENSSON J. A. L., Sweden
ŚWIĄTEK Jerzy, Poland*
SVIREZHEV Yu. M., USSR
SYCH Zbigniew, Poland
SYCHOWICZ Andrzej, Poland*
SYCZEWSKA E. M., Poland
SZABÓ Árpád, Hungary
SZABÓ Zoltán, Hungary
SZAFIRSKI Bolesław, Poland
SZAFRANIEC F. H., Poland
SZAFRANIEC Zbigniew, Poland
SZAL Barbara, Poland
SZALAY Michael, Hungary
SZAŁAJKO Kazimierz, Poland
SZAMKOŁOWICZ Lucjan, Poland
SZAPIRO Tomasz, Poland
SZATKOWSKI Mirosław, Poland*
SZCZEPANEK Wojciech, Poland
SZCZEPANIAK Jerzy, Poland
SZCZEPANKIEWICZ Eugeniusz, Poland
SZCZEPAŃSKI Andrzej, Poland
SZCZERBICKA-LIPIŃSKA Helena, Poland
SZCZĘSNY Wiesław, Poland
SZCZOTKA Franciszek A., Poland
SZCZYPIŃSKI Tomasz, Poland
SZELMECZKA Jerzy, Poland
SZENTMIKLÓSSY Zoltán, Hungary
SZEPIETOWSKI Andrzej, Poland*
SZILASI József, Hungary
SZLENK Wiesław, Poland
SZMANDA Błażej, Poland
SZMIDT Janusz, Poland
SZMYDT Zofia, Poland
SZNAJDER Roman, Poland
SZÖKEFALVI-NAGY Béla, Hungary

SZOSTEK Grzegorz, Poland
SZPIRGLAS Jacques, France
SZPIRGLAS Viviane, France
SZTAJNIC Jerzy, Poland
SZULGA Jerzy, Poland
SZULKIN A., Sweden
SZUREK Michał, Poland
SZWARC Ryszard, Poland
SZYMANIAK Piotr, Poland
SZYMAŃSKI A. A., Poland
SZYMAŃSKI Bogusław, Poland
SZYMAŃSKI Wacław, Poland
SZYMICZEK Kazimierz, Poland
SZYSZKOWICZ Mieczysław, Poland

TABISZ Krzysztof, Poland
TABOW Jordan, Bulgaria
TAGAMLITZKI Yaroslav, Bulgaria
TAJIMA Shinich, Japan
TAKAHASHI Kuniharu, Japan
TAKHTAJAN L. A., USSR
TALLINI Giuseppe, Italy
TALLINI SCAFATI Maria, Italy
TAM B. S., China-Taiwan
TAMÁSSY Lajos, Hungary
TAMBURINI Maria-Clara, Italy
TAMM Martin C. Å., Sweden
TAMRAZOV P. M., USSR
TARJAN R. E., USA
TARKOWSKI Stanisław, Poland
TATARKIEWICZ Krzysztof, Poland
TAVAKOLI Javad, Iran
TCHAKERIAN K. B., Bulgaria
TEISSIER Bernard, France
TEKLEHAMANOT Retta, Ethiopia
TELESNIN V. R., USSR
TELJAKOVSKIĬ S. A., USSR
TEMBROWSKI Bronisław, Poland
TEMLJAKOV V. N., USSR
TEMPELMAN Arkady A., USSR
TEMPERLEY D. J., UK
TERADA Fumiyuki, Japan*
TERENTIEV Evgeniĭ D., USSR
TER-KRIKOROV A. M., USSR
TETERIN Yu. G., USSR
THEIS Antoinette, France
THIELE H., GDR
THOLEN W. P., FRG
THOMPSON A. C., Canada
THOMPSON T. M., USA
THURSTON William, USA
TICHÁ Marie, Czechoslovakia
TIIT Ene-Margit A., USSR
TIMOKHOV Aleksandr S., USSR
TITZOVA Nadezhda, Bulgaria
TOBIAS T. A., USSR
TOCZYŁOWSKI E., Poland
TODOROV Kaltscho, Bulgaria
TOGARI Yoshio, Japan
TOLLIS Theodore, Greece
TOLSTED Elmer B., USA
TOMANOV G. M., Bulgaria
TOMIDA Masamichi, Japan
TOMINAGA Akira, Japan
TOMIYAMA J., Japan
TOMSKIĬ G. V., USSR
TONCHEV Vladimir D., Bulgaria
TONEV Toma V., Bulgaria
TONKOV Tonko, Bulgaria
TOPP Jerzy, Poland
TOPPILA Sakari, Finland
TOPSØE F., Denmark
TORUŃCZYK Henryk, Poland
TOURÉ Saliou, Ivory Coast
TOVAR SANCHEZ Luis Manuel, Mexico
TRACZYK P., Poland
TRACZYK Tadeusz, Poland
TRANAH D. A., UK
TRAORE Sekou, Congo
TRAPLE J., Poland
TRAUTMAN Andrzej, Poland
TRELIŃSKI Gustaw, Poland
TRENOGIN V. A., USSR
TRIEBEL Hans, GDR
TRIMECHE Khalifa, Tunesia
TROCHANOWSKI Włodzimierz, Poland
TROTMAN D. J. A., France
TROYANSKI Stanimir, Bulgaria
TRPENOVSKI Branko, Yugoslavia
TRUSZCZYŃSKI Mirosław, Poland
TRZASKALIK Tadeusz, Poland
TRZEPIZUR Andrzej, Poland
TSUBOI Takashi, Japan
TSUBOTA Etsuko, Japan
TSUKADA H., Japan

TURAEV V. G., USSR
TURDZA Erwin, Poland
TURNAU Stefan, Poland
TURNER Nura D., USA
TURO Jan, Poland
TURPIN Ph., France
TURZAŃSKI Marian, Poland
TUTSCHKE W., GDR
TVERBERG H., Norway
TU VU CONG, Viet-Nam
TWARDOWSKA K., Poland
TYC Andrzej, Poland
TYNYANSKIĬ N. T., USSR

UCHIYAMA Saburo, Japan
UDRIŞTE Constantin, Romania
UGOWSKI Henryk, Poland
UHRIN Béla, Hungary
UKHOBOTOV V. I., USSR
UL'YANOV Peter L., USSR
URBANIK Kazimierz, Poland
URBAŃSKA Aleksandra, Poland
URBAŃSKI Mariusz, Poland
URBAŃSKI Ryszard, Poland
USCKI Mirosław, Poland
USTINA Fred, Canada

VAINIKKO Gennadi M., USSR
VÄISÄLÄ Jussi, Finland
VAITERSIĆ Marian, Czechoslovakia
VALEV Vencheslav, Bulgaria
VALLÉE Robert, France*
VAN DEN DRIES L., Netherlands
VAN DER WALT A. P. J., South Africa
VANDEV Dimitar, Bulgaria
VAN GEEL Jan M-H., Belgium
VAN LINT J. H., Netherlands
VAN MOERBEKE Pierre, Belgium/USA
VARCHENKO A. N., USSR
VASIL'EV S. N., USSR
VASILIEVA Radka, Bulgaria
VAZ FERREIRA Arturo, Portugal*
VELIKIN V. L., USSR
VELIOV V. M., Bulgaria
VENKOV A. B., USSR
VERES Sándor, Hungary
VERGNE Michèle, France
VERNESCU B., Romania

VESELÝ Jiři, Czechoslovakia
VERSTERGAARD P. D., Denmark
VESZTERGOMBI Katalin, Hungary
VETULANI Z., Poland
VIGNERAS Marie-France, France
VILIMOVSKÝ Jiři, Czechoslovakia
VILLANI V., Italy
VINCZE Endre, Hungary
VINOGRADOV Stanislav A., USSR
VINOGRADOV Vladimir S., USSR
VITUSHKIN A. G., USSR
VITUSHKINA V. P., USSR
VIVIENTE José Luis, Spain
VLADIMIROV V. S., USSR
VOGEL Wolfgang, GDR
VOICULESCU Dan V., Romania
VOIGT B., FRG
VO KHAC Khoan, France
VOLKMANN Peter, FRG
VOLODIN I. N., USSR
VOLOVICH I. V., USSR
VRKOĆ Ivo, Czechoslovakia
VUORINEN Matti, Finland*

WAADELAND H., Norway
WACHNICKI Eugeniusz, Poland
WACHTA Krystyna, Poland*
WADA Hidekazu, Japan
WAGNER-BOJAKOWSKA Elżbieta, Poland
WAJLER Stanisław, Poland
WAJNBERG Irena, Poland
WAJNRYB Bronislaw, Israel
WAKULICZ Andrzej, Poland
WAKULICZ Antoni, Poland
WALAT Andrzej, Poland
WALCZAK P. G., Poland
WALCZAK Stanisław, Poland
WALDSPURGER Jean-Loup, France
WALIGÓRSKI Stanisław, Poland
WALISZEWSKI Włodzimierz, Poland
WALKER M. F., South Africa
WALL G. E., Australia
WALSCH Werner, GDR
WALUKIEWICZ Henryk, Poland
WANG Shu-Tang, China*
WARZECHA-SAWOŃ Aleksandra, Poland

WASCHBÜSCH Josef, FRG
WAŚKO Anna, Poland
WASZAK Aleksander, Poland
WATANABE Hiroshi, Japan
WATANABE Kimio, Japan
WATANABE Shinzo, Japan
WATARI Nozomu, Japan
WATSON Saleem, USA
WATTERS John F., UK
WEBER Karl-Heinz, GDR
WECHLER W., GDR
WĘDRYCHOWICZ Stanisław, Poland
WĘGLOWSKI Zbigniew, Poland
WEILL G. G., France/USA
WEINZWEIG A. I., Canada
WEISS Tomasz, Poland
WERBOWSKI Jarosław, Poland
WETTE Eduard W., FRG
WEYMAN Jerzy, Poland
WHEELER D. H., Canada
WHITELEY Walter, Canada
WIDOM Harold, USA
WIELECHOWSKI Grzegorz, Poland
WIERMAN John C., USA
WIERZBICKA Danuta, Poland
WIGGINS K., USA
WILCZYŃSKI Władysław, Poland
WILDENAUER Peter, FRG*
WILDENHEIN Günter, GDR
WILKINS David R., UK
WILSON J. S., UK
WINIARSKA Teresa, Poland
WINIARSKI Tadeusz, Poland
WINIARZ Adam, Poland
WIŚNIEWSKI Henryk, Poland
WIŚNIEWSKI Jarosław, Poland
WIŚNIEWSKI Kazimierz, Poland
WISTUBA Irena, Poland
WITKOWSKI Alfred, Poland
WITOMSKI Jerzy, Poland
WITTMAN Rainer, FRG*
WIWEGER Antoni, Poland
WNUK Witold, Poland
WOJCIECHOWSKA Agnieszka, Poland
WOJCIECHOWSKI Krzysztof, Poland
WÓJCIK A. P., Poland
WÓJCIK A. Ryszard, Poland
WÓJCIK Jan, Poland
WOJDYŁO Bolesław, Poland
WOJTASZCZYK Przemysław, Poland
WOJTCZAK Leszek, Poland
WOJTOWICZ Jacek, Poland
WÓJTOWICZ Zygmunt, Poland
WOJTYŃSKI Wojciech, Poland
WOLAK Robert A., Poland
WOLD A. L., Sweden
WOLNIEWICZ Tomasz, Poland
WOŁODŹKO Stanisław, Poland
WOLSKA-BOCHENEK Janina, Poland
WONG Bun, Hong-Kong
WONG P. M., Hong-Kong
WOOLF William B., USA
WORONOWICZ Stanisław, Poland
WOŚ Janusz, Poland
WOWK Czesław, Poland
WOŹNIAKOWSKI Henryk, Poland
WRĄBEL Bogusław, Poland
WRÓBEL Katarzyna, Poland
WRONA Włodzimierz S., USA
WRONICZ Zygmunt, Poland
WROŃSKA Ewa, Poland
WROŃSKI S., Poland
WUCZYŃSKA Krystyna, Poland
WULFSOHN Aubrey, South Africa
WUSSING H., GDR
WYDERKA Zdzisław, Poland
WYLER Armand, Switzerland*
WYRWIŃSKA Aleksandra, Poland

XIE Bang-gie, China*

YAFAEV D. R., USSR
YAKOVLEV A. V., USSR
YANOVICH Leonid A., USSR
YASEEN Adel, Kuwait
YAU Shing-Tung, USA
YAU Stephen S. T., USA
YISMAN Alemu, Ethiopia
YOSHIMOTO Takeshi, Japan
YULMUKHAMETOV R. S., USSR
YURCHUK Nicolai I., USSR

ZAARE-NAHANDI Rahim, Iran
ZABCZYK Jerzy, Poland
ZĄBEK Ryszard, Poland

ZABOLSKI Tomasz, Poland
ZACHAREK Stanisław, Poland
ZACHARIAS G., USA
ZACHARIOU Andreas, Cyprus
ZAFARANI Jafar, Iran
ZÄHLE M., GDR
ZAHORSKI Zygmunt, Poland
ZAJĄCZKOWSKA Maria, Poland
ZAJĄCZKOWSKI Wojciech, Poland
ZAJTZ Andrzej, Poland
ZAK F. L., USSR
ŻAK Tomasz, Poland
ZAKHAROV V. E., USSR
ŻAKOWSKI Wojciech, Poland
ZAKRZEWSKI Piotr, Poland
ZALEWSKI Romuald, Poland
ZALEWSKI Zbigniew, Poland
ZAMEK-GLISZCZYŃSKI Tomasz, Poland
ZAMYATIN A. P., USSR
ZANDECKA-PERL Urszula, Poland
ZAPPA Guido, Italy
ZARACH Andrzej, Poland
ZAREMBA L. S., Poland
ZAREMBA Piotr, Poland
ZARZYCKI Piotr, Poland
ZASTAWNIAK Tomasz, Poland
ZAWADOWSKI W. M., Poland
ZBAGANU Gheorghita, Romania
ZDUN Marek, Poland
ZDUNIK Anna, Poland
ZEIDLER E., GDR
ŻEKANOWSKI Zbigniew, Poland
ŻELAZKO Wiesław, Poland
ZELEZNIKOV John, USA
ZEL'MANOV Efim I., USSR
ZEMÁNEK Jaroslav, Poland
ZHIDKOV E. P., USSR
ZHIZHCHENKO A. B., USSR
ZHOU Bo-Xun, China*
ZHOU Xue-Guang, China*
ZHUK I. K., USSR
ZHURBENKO I. G., USSR
ZIAPKOV Nicola, Bulgaria
ZIELIŃSKI Andrzej, Poland
ZIELIŃSKI Ryszard, Poland
ZIELIŃSKI Wojciech, Poland
ZIEMIAN Bogdan, Poland
ZIEMIAN Krystyna, Poland
ZIEMIŃSKA Jolanta, Poland
ZIĘTAK Krystyna, Poland
ZIL'BER B. I., USSR
ZINK Th., GDR
ZIPPO Shiro, Japan
ZIRILLI Francesco, Italy
ZIZLER Vaclav, Czechoslovakia*
ZNÁM Stefan, Czechoslovakia
ZOGRAF Peter, G., USSR

ŻOŁĄDEK Henryk, Poland
ŻOŁNOWSKA Halina, Poland
ZOLOTAREV V. M., USSR
ZORSKI Henryk, Poland
ZUBAREV D. N., USSR
ŻUCHOWSKI Ryszard, Poland
ZWIERZEWICZ Zenon, Poland
ZYCH Edward, Poland
ZYGMUND Antoni, USA
ŻYWIEŃ Łucja, Poland

Opening Ceremonies

The opening ceremonies of the Warsaw Congress took place in the Congress Hall of the Palace of Culture on August 16, 1983 at 9.30. After a performance of the men's choir of the Choral Society "Harfa" (The Harp), which started with the national anthem, Professor Lennart Carleson, President of the International Mathematical Union for the years 1979–1982 opened the proceedings with the following words:

> On behalf of the International Mathematical Union I am happy to greet you all here today to begin the work of the 1982 International Congress of Mathematicians. Already at the Zürich meeting in 1897 it was stated that the first objectives of the Congress are to promote the personal relations between mathematicians from different countries and to give a survey of the state of our science. The rules of the congresses have, through the years, become firmer and since 1962 the IMU is formally responsible for the scientific content. This Congress meets under special circumstances, but the main objectives remain and to keep unbroken traditions has been a fundamental concern to the IMU.
>
> The organization of the Congress is by our rules in the hands of an organizing committee. Following a well established tradition I now propose that the president of the organizing committee, Professor Czesław Olech, is elected President of the Warsaw Congress.

The proposal of Professor Carleson was accepted by acclamation. Following his election Professor Olech gave his presidential address to the Congress.

> It is my pleasant duty to declare the 1982 International Congress of Mathematicians open.
>
> This is a great and happy moment for the Polish mathematical community, strongly represented here, on whose behalf I would like to welcome cordially all foreign participants.

On behalf of all of you, I would like to welcome the President of the Polish Academy of Sciences, Professor Aleksander Gieysztor. We greatly appreciate the fact that he consented to be the Patron of the Warsaw Congress and that he has been so kind as to attend this ceremony in person.

I also extend a most cordial welcome to all our distinguished guests. Among them, I wish to greet Professor Zdzisław Kaczmarek, the Scientific Secretary of the Polish Academy of Sciences, Professor Stanisław Nowacki, Deputy Minister of Science, Higher Education and Technology, Mr Józef Wiejacz, Deputy Minister of Foreign Affairs, Mr Stanisław Szewczyk, Vice-President of the City of Warsaw.

Organizing a meeting of this magnitude would not have been possible without the active support of the Government.

At an early stage, before the International Mathematical Union (IMU) had made its final decision, the Secretary of the Polish Academy of Sciences explicitly promised to back the Congress should it be held in Poland; that promise has been fulfilled in all aspects. We have likewise received the support of the various governmental agencies throughout the preparations for the Congress. For the support and smooth cooperation of these agencies, I would like to express thanks and appreciation.

Professor Urbanik, inviting you to Warsaw at the closing ceremony in Helsinki, said: "For a long time Polish mathematicians have carried in their hearts the desire to organize an international Congress".

The best example of those was Professor Kazimierz Kuratowski, once very active in the IMU, who was strongly advocating the initiative to invite the Congress to Poland. I am very sorry that he did not live long enough to share with us this happy moment.

The hope that the desire to have the Congress in Poland would become a reality was based on the belief that the rich tradition of mathematical research carried out in this country makes Poland an acceptable choice for the Site Committee.

We are privileged to have with us Professor Władysław Orlicz, the Nestor of Polish mathematicians, who, for more than fifty years has been enriching this tradition in Poland.

I propose that Professor Orlicz be elected Honorary President of the Congress.

The proposal was warmly accepted by the Congress and Professor Władysław Orlicz was elected Honorary President by acclamation.

The general aim of an ICM is to give an appraisal of current mathematical research. This important and difficult task could not be reached without wide international cooperation, and the active involvement and hard work of many leading mathematicians.

The official mathematical programme is decided upon by an international Consultative Committee established for ICM by IMU. This Committee, after two years' work, produces the final list of invited speakers, taking into account the proposals of the panels and the suggestions of the National Committees.

The Consultative Committee for the present Congress consists of six members appointed by IMU — Professor Jean-Pierre Serre, chairman, and Professors Michael Atiyah, William Browder, Pierre Deligne, Ludvig Faddeev and Shmuel Winograd.

Three other members, Bogdan Bojarski, Zbigniew Ciesielski and Stanisław Łojasiewicz, represent the Organizing Committee.

May I propose that we thank the Consultative Committee for the work they have done and that we also extend our thanks to all those involved in preparing the programme.

The organizational responsibility for the Congress was shared by the Department of Mathematics of the Warsaw University, the Institute of Mathematics of the Warsaw Technical University and the Institute of Mathematics of the Polish Academy of Sciences. A number of mathematicians from outside Warsaw were also members of the Organizing Committee.

Many institutions and individuals, mathematicians and members of the administrative staff have been involved in the preparations. I wish to thank them all for their hard work for their support, for sharing with me the responsibility for the Congress.

The role of the Organizing Committee is mainly technical. This time, however, the Organizing Committee was faced with some extra responsibility when the question of holding the Congress in Warsaw was again raised and discussed.

In April 1982, the Executive Committee of IMU, considering the scientific prospects for the ICM-82 at that time, decided to postpone the Warsaw Congress by one year. This decision was accepted by the Organizing Committee in the conviction that it would be advantageous for the final result.

The final scientific result you will be witnessing yourselves. It depends not only on those who have prepared the program but also on the cooperation of those who have been chosen to fill in the programme with the invited survey lectures, both plenary and in sections. I regret that you will not have the opportunity to listen to some of the lecturers announced in the Third Announcement or even in the printed Programme you have just received.

Nevertheless the number of invited speakers present at the Congress, though not full, is over one hundred and I would like to welcome them particularly warmly.

Applying for the Congress in Warsaw, we expected that this would be an opportunity for greater participation in an ICM of mathematicians from Poland and other socialist countries. I would like to observe with great satisfaction that our expectations have become a reality.

Warsaw is a known centre for mathematical research. It was here that the first specialized international journal of mathematics in the world was founded. I am speaking of the Fundamenta Mathematicae. Here, for the last ten years, mathematicians from all over the world meet regularly at the Stefan Banach International Mathematical Center, a common enterprise of the Academies of socialist countries.

Let me express the hope that this Congress will contribute to all these international mathematical traditions to a considerable extent.

Professor Olech left the floor to the Patron of the Congress, Professor Aleksander Gieysztor, President of the Polish Academy of Sciences, who delivered the following address to the Congress:

Mr Honorary President, Mr President of the Congress, Ladies and Gentlemen,

It is a great pleasure and honour for me to welcome all of you on behalf of the Polish Academy of Sciences, the scholarly community of this country and the city of Warsaw. Our Academy from the very moment when the idea of holding the International Mathematical Congress in Warsaw appeared, declared its support to our colleagues involved in its organization. The idea of having the Congress in Warsaw was in the air for quite some time and was especially cherished by the late Kazimierz Kuratowski, Vice-President of our Academy and a scholar who contributed greatly to the organization of Polish mathematics during the past several decades.

The decision of the International Mathematical Union in 1978 to have the Congress in Warsaw undoubtedly expressed an appreciation for the contribution of Polish mathematics to the general body of mathematical knowledge. I assume this appreciation pertains both to past accomplishments and to the current state of mathematical research in Poland.

The difficulties in preparing the Congress have been many and of a quite varied nature. It is fortunate for us that most of them have been overcome. Now you are here in Warsaw as representatives of over sixty countries from all over the world, the total number of about two thousand three hundred mathematicians brought together to discuss mathematics and to develop academic and personal contacts. Your presence here is a proof that the idea of international scientific cooperation is strong enough to overcome impediments of any kind.

The months of August and September encompass two important dates in the history of this country. Thirty nine years ago on the 1st of August the Warsaw Uprising began and September 1st 1939 was the first day of the Second World War. Both these months are times of national remebrance, of reflection upon the history of our country.

During the Second World War the Polish scientific community was decimated. In particular, well over half of the actively working Polish mathematicians lost their lives. Many others found themselves in various countries all over the world. Universities, libraries and printing presses in Poland were largely destroyed. The educational system of the country was in ruins and scientific activity was disrupted.

The fact that this Congress is being held in Warsaw in 1983, thirty eight years after the war, gives evidence of the reconstruction of Polish science both in the organizational and the substantive sense. In particular, it is a proof of the renaissance and expansion of the Polish mathematical community. At present this community is many times larger than before the Second World War. The membership of the Polish

Mathematical Society has increased at least fifteenfold and there are active scientific groups in many universities and polytechnics. The Institute of Mathematics of the Polish Academy of Sciences has a very well supplied mathematical library, which serves the whole of Polish mathematics. This year we have celebrated the 10th anniversary of the activities of the Stefan Banach Mathematical Center.

In these facts one can see a telling example of the vitality and enduring power of the cultural and intellectual heritage of mankind. On the other hand, this heritage has to be constantly regenerated with new creative forces.

Science is not a simple continuative accumulation of knowledge, but changes of level and crises. The levels are formed by the normal evolution of science, by the solution of particular problems according to steady and unquestionnable paradigms. The crises are scientific revolutions, changes of paradigms.

It is remarkable to observe that mathematics, this oldest science, together with logic, is constantly generating new concepts, creating new domains of mathematical research, embracing and influencing ever wider areas of human knowledge and understanding of the world in general.

In the history of human thought we can easily trace many examples of principal importance showing how mathematical thinking essentially contributed to the fundamental change of our basic concepts of the structure of natural phenomena. We have only to think about the ideas of Copernicus, the relativity theory of Einstein or quantum mechanics. At present we are witnessing the impact on our lives and ways of thinking of the scientific and technological revolution based on the incredible possibilities of computers and their overall invasion. The resulting changes in our civilization are not comparable in their scope and rapidity to anything experienced in the past. The social, economic, cultural and political implications of these changes are enormous.

What is the driving force of this development? What is it that keeps this wonderful edifice growing, integrated and relevant for the understanding of the world around us and having an esthetic value as well?

Why should this rapidly expanding area of knowledge not have broken apart into separate and disconnected branches?

All these facts, questions, perhaps even answers, you understand much better than anyone outside your science. Of course I don't dare to answer these questions. I would rather listen. And reading the words of John von Neumann: "It is undeniable that some of the best inspirations in mathematics — in those parts of it which are as pure mathematics as one can imagine — have come from the natural sciences" and those of Aleksandr Danilovich Aleksandrov stating that "...the vitality of mathematics arises from the fact that its concepts and results, for all their abstractness, originate... in the actual world", I see more clearly the dual character of the relationship between mathematics and science in general. Mathematics is the leafage on the tree of science and contributes to the welfare of the whole structure but in order to live it must essentially depend on its roots.

Let me recall at this moment what David Hilbert said: "The organic unity of mathematics is inherent in the nature of this science, for mathematics is the foundation of all exact knowledge of natural phenomena. That it may completely fulfil this high mission, may the new century bring it gifted masters and many zealous and enthusiastic disciples."

The fact that you have systematically organized your congresses since 1900 testifies that the idea of Hilbert about the unity of mathematics is as valid now as it was eighty three years ago.

The fact that so many young people engage in mathematics and attend the Congresses and that to the most outstanding of them you award the golden medals of international recognition shows that you have succeeded in attracting "enthusiastic disciples". I know that you make conscious and systematic efforts to recruit and shape these disciples. After all, it is in mathematics that the idea of Olympic Competition for high school students arose. We note with pleasure the activities of the International Commission on Mathematical Instruction accompanying the Mathematical Congress. You care for the young generation and you also care for outsiders to your science. You understand very well that all these efforts are in the long run essential for the vigour of your science, for its ability to grow and yet remain young.

I believe these actions and the attitude giving rise to them are the right ones in the much more general setting of "la Condition Humaine". Perhaps the mystery of staying young and of maintaining permanent development is revealed in these words: give your best, take in selectively.

I wish all of you and your fascinating science perseverance in pursuit of these goals.

I assume that, because of the universality and wide applicability of mathematical thinking, this Congress will be also relevant for science and human culture in general.

I would like to express my hope that the several days you spend in Poland will allow you to see a little of our country and give you an opportunity for direct encounters with Polish people, leading to a better understanding of our thinking, history and current problems. Let me express my best wishes for the success of the Congress.

Professor Carleson, Chairman of the Fields Medals Committee, then presented the following report:

One of the most important activities of each congress is the award of the Fields prizes. The list of mathematicians who through the years have received these prizes is indeed impressive. Work in the committee for these awards gives an even stronger impression of the strength and breadth of current mathematical research and of the vitality of the young generation of mathematicians.

The idea of J. C. Fields to encourage young mathematicians has been a great success. As we are all aware, our science is now closely tied to the revolution of society created by computers. For this reason the IMU has accepted with great satisfaction an offer by the University of Helsinki, Finland, to finance a prize in "Mathematical aspects of Information Science" with objectives similar to those of the Fields prizes. In recognition of Rolf Nevanlinna's contribution to our science, both in the IMU and in Finnish computer science, the prize has been named the Nevanlinna prize.

The committee for the Fields prize had the following members: H. Araki, Kyoto; N. Bogolyubov, Moscow; P. Malliavin, Paris; D. Mumford, Cambridge, Mass.; L. Nirenberg, New York; A. Schinzel, Warsaw; C. T. C. Wall, Liverpool, and myself as chairman.

The committee for the Nevanlinna prize consisted of: J. L. Lions, chairman, J. Schwartz, New York and A. Salomaa, Turku.

The decisions of the committees were already announced at the meeting of IMU in August 1982 here in Warsaw. The Fields prizes were awarded to Alain Connes, Paris; William Thurston, Princeton and Shing-Tung Yau, Princeton — and the Nevanlinna prize to Robert Tarjan, Stanford. I offer them our warmest congratulations.

May I now ask our Honorary President, Professor Orlicz, to come forward and give the prizes to the winners.

When Professors Connes, Tarjan, Thurston and Yau had received the prizes from Professor Orlicz, it was announced that after the opening session Professor H. Araki would speak on the work of Connes and Professor Schwartz on that of Tarjan, and that the reports of Professor Wall on the work of Thurston and of Professor Nirenberg on the work of Yau will be read out in their absence.

The opening session ended with a concert of the men's choir "Harfa" conducted by Michał Dąbrowski.

Closing Ceremonies

The closing session of the Warsaw Congress took place in Congress Hall to the Palace of Culture on August 24, 1983 at 12.15. Professor Olli Lehto, fefl Secretary of the International Mathematical Union, delivered the lohowing address:

Ladies and gentlemen, dear colleagues,

It has been customary at the closing session of the Congress to present a report on the activities of the previous General Assembly of the International Mathematical Union. Usually, the General Assembly is held just before the Congress, but this time, the 1982 General Assembly took place already a year ago. It was held here in Warsaw and was attended by the delegates of almost all member countries of IMU. With some experience from previous assemblies, I found that the Warsaw meeting had an exceptionally friendly atmosphere.

The resolutions of this General Assembly have been published in an issue of the IMU Bulletin which has been distributed long ago. Therefore, I think there is no reason to go into details here now.

As you may well guess, much of the discussion at the General Assembly centered around the ICM-82. The decision to hold the 1982 Congress in Warsaw was made by the IMU Site Committee in Finland in 1978. It was regarded as a very good decision, justified by the fine mathematical tradition in Poland and by the fact that Poland seemed to be easily accessible to mathematicians no matter which part of the world they came from. For three years, the organization of the Congress was running smoothly. But as we all know, difficulties started at the end of 1981, just at the time when the main pre-Congress document, the Second Announcement, was being sent out.

The changed situation put the mathematicians to a waiting position, as it was not known whether the Congress could be held. It soon became clear that, in spite of the new conditions, the Poles were willing to continue their efforts for organizing the Congress. However, in April 1982 the Executive Committee of IMU arrived at the conclusion that conditions for a scientifically good congress did not exist in August 1982. Making the final decision about the Congress was postponed till November. This procedure was endorsed by the General Assembly, which also provided the Executive Committee with much useful advice.

It was not an easy task for the Executive Committee to reach a decision. But after weighing the pros and contras, the Executive Committee unanimously decided that the ICM-82 be held in Warsaw in August 1983.

It is of course not up to me to make a general evaluation in public of whether our decision was correct or not. But let me say that I feel very happy that the ICM-82 now took place here. The continuity of international cooperation was maintained, and in spite of regrettable absences of some invited speakers, this was a high class meeting from the scientific point of view.

The positive feeling towards this Congress is also largely due to the excellent work done by our Polish colleagues. We have all seen how well everything is functioning, and we have sensed the warm and friendly atmosphere of the Congress. All this can only be achieved by the joint effort of a large number of people. Our thanks are due to all of them, the more so, as the work has been carried out under such difficult circumstances.

An exceptionally heavy load has been on the shoulders of one person, the chairman of the Organizing Committee and President of ICM-82, Professor Czesław Olech. His skill and strength have largely contributed to the success of this Congress.

At the time of the General Assembly, we did not know whether ICM-82 would be held. Nevertheless, it was then already time to think of the 1986 Congress. The General Assembly confirmed the decision of the IMU Site Committee to accept the invitation of the United States National Academy of Sciences to ICM-86 at the University of California, Berkeley.

The IMU has a Special Development Fund whose principal aim is to help young mathematicians from developing countries

> to take part in ICM's. This time, the Union was able to give grants to 33 mathematicians from 21 different countries. Since the funds of the Union are limited, the success of this important project depends largely on donations made to the Special Development Fund. Fund raising for ICM-86 has already started, in that an appeal was recently sent to the National Committees for Mathematics of the member countries of IMU.
>
> Let me conclude by emphasizing the basic principle of IMU that politics should never find a foothold within the Union. As individuals, we may of course have whatever political views we choose, but when it amounts to organized international cooperation in mathematics, then political aspects should be put aside entirely. Our fine science should be the uniting link between us and make us in a true sense one big mathematical family.

Professor O. Lehto then invited Professor Jack K. Hale to speak on behalf of the Mathematical Community of the United States of America. Professor Hale spoke as follows:

> We are very pleased that the next ICM will be held in Berkeley, California, USA.
>
> On behalf of the mathematical community of the United States, I would like to extend a cordial invitation to all of you to attend that congress. My only hope is that we can be as gracious a host as our colleagues in Poland and that the congress will be as well organized and as successful as the ICM in Warsaw.

The invitation was accepted by acclamation.

Speaking on behalf of the members of the Congress, Professor Hans Freudenthal and Academician S. M. Nikol'skiĭ expressed their thanks to their Polish hosts.

In his reply, Professor Olech thanked the speakers for their warm words of appreciation of the work of the organizers of the Warsaw Congress. He gave some statistical information about the Congress, thanked the Congress members for their contributions, particularly all the speakers and the chairmen of the various sessions. He passed on the words of thanks to all his Polish colleagues who participated in the arrangements and to the members of the Congress Bureau. He closed his address as follows:

> In our work we have used much of the experience of the organizers of the previous Congress in Helsinki. In many cases we followed exactly the procedure they used. This greatly simplified our work and was a great help for us. I was in constant contact with Professor Olli Lehto, the President of the Helsinki Congress, and, in particular, his personal advice was of great value. I wish to thank him for all that very much and ask him to transmit our thanks to our Finnish colleagues. If we can pay this back by being of any use to the organizers of the next Congress in Berkeley, we shall be only too happy.
>
> Let me thank you all for coming to the Congress in Warsaw. I hope you have enjoyed your stay both mathematically and socially. I wish all of you all the best for the years to come.
>
> This way we came to what I personally consider a happy end and I declare the ICM-82 in Warsaw closed.

The Work of the Medallists

The work of the Fields medalists and the Nevanlinna prize winner were presented as follows (the asterisk means that the report was read out in the absence of the author).

Fields Medallists

Hozihiro Araki — The Work of Alain Connes
C. T. C. Wall* — The Work of W. Thurston
Louis Nirenberg* — The Work of Yau, Shing-Tung

The Nevanlinna prize winner

J. Schwartz — The Work of Robert Endre Tarjan

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

HUZIHIRO ARAKI

The Work of Alain Connes

Theory of operator algebras, after being quietly nourished in somewhat isolation for 30 years or so, started a revolutionary development around late 1960's. Alain Connes came into this field just when the smokes of the first stage of the revolution were settling down. He immediately led the field to breathtaking achievements beyond the expectation of experts.

His most remarkable contributions are: (1) general classification and a structure theorem for factors of type III, obtained in his thesis [12], (2) classification of automorphisms of the hyperfinite factor [29], which served as a preparation for the next contribution, (3) classification of injective factors [31], and (4) applications of the theory of C^*-algebras to foliations and differential geometry in general [44, 50] — a subject currently attracting a lot of attention.

In this report, I shall mostly concentrate on the first three aspects which form a well-established and most spectacular part of the theory of von Neumann algebras.

1. Classification of type III factors

In the latter half of 1930's, Murray and von Neumann initiated the study of what are now called von Neumann algebras (i.e. weakly closed *-subalgebra of the *-algebra $L(H)$ of all bounded linear operators on a Hilbert space H) and classified the factors (i.e. von Neumann algebras with trivial centers) into the types I_n, $n = 1, 2, \ldots$, and I_∞ (isomorphic to $L(H)$ with $\dim H = n$ and ∞), II_1, II_∞ and III. (In the following we restrict our attention to von Neumann algebras M on separable Hilbert spaces H.) Only three type III factors (and only three type II_1 factors) had been known to be mutually non-isomorphic till 1967, when Powers showed the existence of a continuous family of mutually non-isomorphic type III factors.

Traces provided a tool for a systematic analysis of type II factors at an earlier stage, while the non-existence of traces made type III factors remain untractable till late 1960's, when the Tomita–Takesaki theory was created and furnished a powerful tool for type III. To introduce notation, let M be a von Neumann algebra on a separable Hilbert space H and let $\Psi \in H$ be cyclic (i.e. $M\Psi$ be dense in H) and separating (i.e. such that $x\Psi = 0$ for $x \in M$ implies $x = 0$). The conjugate linear operator $S_\Psi x\Psi \equiv x^*\Psi$, $x \in M$, has a closure $\bar{S}_\Psi$ and defines the positive selfadjoint operator $\Delta_\Psi \equiv S_\Psi^* \bar{S}_\Psi$, called the *modular operator*. The Tomita–Takesaki theory says that $x \in M$ implies $\sigma_t^\psi(x) \equiv \Delta_\Psi^{it} x \Delta_\Psi^{-it} \in M$. The one-parameter group of (*-)automorphisms σ_t^ψ of M depends only on the positive linear functional $\psi(x) \equiv (x\Psi, \Psi)$ and is called the *group of modular automorphisms*.

Connes [12] has shown that the modular automorphisms for different ψ's are mutually related by inner automorphisms (in other words, the independence of σ_t^ψ from ψ in the quotient $\operatorname{Out} M$ of the group $\operatorname{Aut} M$ of all automorphisms modulo the subgroup $\operatorname{Int} M$ of all inner automorphisms) and introduced the following two isomorphism invariants for M:

$$S(M) = \bigcap_{\Psi} \operatorname{Sp} \Delta_\Psi \quad \text{(Sp denotes the spectrum)},$$

$$T(M) = \{t \in \boldsymbol{R} \colon \sigma_t^\psi \in \operatorname{Int} M\}.$$

(In a more general case, $S(M) = \bigcap \operatorname{Sp} \Delta_\psi$, where the intersection is taken over faithful normal semifinite weights ψ.) It turns out that $S(M) \setminus \{0\}$ is a closed multiplicative subgroup of $\boldsymbol{R}_+^*$, and this leads to the classification of type III factors into the types III_λ, $0 \leqslant \lambda \leqslant 1$:

$$S(M) = \{\lambda^n \colon n \in \boldsymbol{Z}\} \cup \{0\} \quad \text{if } 0 < \lambda < 1,$$

$$S(M) = \boldsymbol{R}_+^* \quad \text{if } \lambda = 1, \qquad S(M) = \{0, 1\} \quad \text{if } \lambda = 0.$$

The Powers examples R_λ, $0 < \lambda < 1$, due to Powers, are of types III_λ and hence mutually non-isomorphic. The two invariants $r_\infty(M)$ and $\varrho(M)$ of Araki and Woods, introduced for a systematic classification of infinite tensor products of type I factors (including R_λ), are shown to be equivalent to $S(M)$ and $T(M)$ for them [7].

2. Structure analysis of type III factors

Connes [12] went on and succeeded in analysis of the structure of type III_λ factors M, $0 \leqslant \lambda < 1$, in terms of a type II von Neumann algebra N (with a non-trivial center) and an automorphism θ of N, such that M is the so-called *crossed product* $N \times_\theta \boldsymbol{Z}$ of N by θ.

For $0 < \lambda < 1$, θ should scale a trace τ of a type II_∞ factor N in the sense that $\tau(\theta x) = \lambda\tau(x)$ for all $x \in N_+$ and $N_i \times_{\theta_i} \boldsymbol{Z}$, $i = 1, 2$, are isomorphic if and only if there exists an isomorphism π of N_1 onto N_2 such that $\pi^{-1}\theta_2\pi\theta_1^{-1}$ is inner, or equivalently (in view of a later result of Connes and Takesaki), $\pi\theta_1\pi^{-1} = \theta_2$. This means that the pair (N, θ) is uniquely determined by M and the classification of M is reduced to that of the pair (N, θ).

For $\lambda = 0$, θ should scale down a trace τ of a type II_∞ von Neumann algebra N in the sense that $\tau(\theta x) \leqslant \varrho\tau(x)$ for all $x \in N_+$ for some $\varrho < 1$ and, again, there is a somewhat more complicated uniqueness result for the pair (N, θ).

Motivated by the above results of Connes, a general structure theorem including the type III_1 has been obtained by Takesaki in terms of a one-parameter group θ_t of trace-scaling automorphisms of a type II_∞ von Neumann algebra N.

In the process of developing the above classification and structure theory, Connes introduced two important technical tools, namely the unitary Radon–Nikodym cocycle (equivalently, relative modular operators) useful in application to quantum statistical mechanics, non-commutative L_p theory, etc., and the Connes spectrum useful in the analysis of C^* dynamical systems.

3. Classification of automorphisms of the hyperfinite factor

A von Neumann algebra, containing an ascending sequence of finite-dimensional subalgebras with a dense union, is called *approximately finite-dimensional* (AFD). AFD factors of type II_1, as shown by Murray and von Neumann, are all isomorphic to what is called the *hyperfinite factor*, denoted by R in the following. Connes [29] has given a complete classification of automorphisms of R modulo inner automorphisms (i.e. the conjugacy class of $\operatorname{Out} R$). Namely, a complete set of isomorphism invariants in $\operatorname{Out} R$ for an $\alpha \in \operatorname{Aut} R$ is given by the pair of the outer period p $(= 2, 3, \ldots)$, which is the smallest $p > 0$ such that α^p is inner, defined to be 0 for outer aperiodic α, and the obstruction γ which is the

p-th root of 1 (1 for $p = 0$) such that $\alpha^p = \mathrm{Ad}\,U$, $\alpha(U) = \gamma U$, where $(\mathrm{Ad}\,U)(x) = UxU^*$. Although the result is about a specific R, this factor R is in the bottom of all known AFD factors and the result that outer aperiodic automorphisms of R are all conjugate up to inner automorphisms is essential for the results described in the next section.

As a by-product, Connes [23] solved negatively one of old problems on von Neumann algebras by exhibiting, for each $0 < \lambda < 1$, factors of type III_λ, not anti-isomorphic to themselves.

4. Classification of AFD factors

A complete classification of AFD factors of type III_λ, $\lambda \neq 1$, is what I consider the most distinguished work of Connes. It turns out that an AFD factor of type III_λ is unique and is isomorphic to R_λ for each $0 < \lambda < 1$, while AFD factors of type III_0 are isomorphic to Krieger's factors associated with single non-singular ergodic transformations of the Lebesgue measure space, their isomorphism classes being in one-to-one correspondence with the metric equivalence classes of non-singular non-transitive ergodic flows on the Lebesgue measure space.

One of the most important technical ingredients of the proof is the equivalence of various concepts about a von Neumann algebra which arose over years in theory of von Neumann algebras. Murray and von Neumann found a factor N of type II_1 non-isomorphic to R (distinguished by Property Γ). In 1962, Schwartz distinguished N, R and $N \otimes R$ by the following Property P: A von Neumann algebra M on a Hilbert space H has the property P iff for any $T \in L(H)$, the norm closed convex hull of the uTu^* with u varying over all unitary operators in M intersects M'. Any AFD factors posseses Property P.

The Property P for M implies the existence of a projection of norm 1 from $L(H)$ to M'. This is the Hakeda–Tomiyama extension property for M', called Property E by Connes, and is stable under taking the intersection of a decreasing family, the weak closure of the union of an ascending family, the commutant, tensor products and crossed product by an amenable group. Thus the Property P implies Property E for M.

Any projection E of norm 1 from a C^*-algebra $\mathfrak{A}_1$ to its subalgebra $\mathfrak{A}_2$ is shown by Tomiyama to be a completely positive map satisfying the property of the conditional expectation: $E(axb) = aE(x)b$ for any $a, b \in \mathfrak{A}_2$ and $x \in \mathfrak{A}_1$. A C^*-algebra $\mathfrak{A}$ with unit is called *injective* if any completely positive unit preserving linear map θ from $\mathfrak{A}$ into another

C^*-algebra $\mathfrak{B}$ with unit has an extension $\bar{\theta}$ to any C^*-algebra $\mathfrak{A}_1$ containing $\mathfrak{A}$ as a completely positive unit preserving linear map from $\mathfrak{A}_1$ into $\mathfrak{B}$. A von Neumann algebra M is injective if and only if it has Property E.

Effros and Lance called a von Neumann algebra M *semidiscrete* if the identity map from M into M is a weak pointwise limit of completely positive maps of finite rank and proved that M is injective if it is semidiscrete.

Connes [31] unified all these concepts by showing that they are all equivalent. The core result is the isomorphism of all injective factors of type II_1 to the unique hyperfinite factor R; it is established by a highly involved and technical proof, utilizing a theorem on tensor products of C^*-algebras, the property Γ of a factor which Murray and von Neumann introduced to distinguish some factors, properties of $\operatorname{Aut} N$ and $\operatorname{Int} N$ of a factor N, the ultra product R^ω for a free ultrafilter ω, an argument analogous to Day–Namioka proof of Følner's characterization of amenable groups, etc.

The uniqueness of injective factors of type II_1 then implies the uniqueness of injective factors of type II_∞. Together with an earlier uniqueness result for trace-scaling automorphisms of $R \otimes B(H)$ (exhibiting the unique injective factor of type II_∞), it also implies the uniqueness of injective factors of type III_λ, $0 < \lambda < 1$. With help of an earlier result of Krieger, injective factors of type III_0 are also completely classified by the isomorphism class of the so-called flow of weight. Thus Connes succeeded in a complete classification of AFD factors (which is as much as saying injective factors) except for the case of type III_1, which still remains open.

The work of Connes also shows that any continuous representation of a separable locally compact group G generates an injective von Neumann algebra if G/G_0 is amenable, where G_0 is the connected component of the identity (in particular, if G is connected or amenable).

5. Other works

After his success in the almost complete classification of injective factors, Connes turned his attention to application of operator algebras to differential geometry. Connes developed a non-commutative integration theory, which provides a method of integration over a family of ergodic orbits or over the set of leaves of a foliation. One significant outcome of this theory is an index theorem for foliation. I am sure that this subject will rapidly develop much further. For survey of the present status, we refer to [44], [50].

The works on positive cones [13] provide a geometric characterization of von Neumann algebras through the associated natural positive cone in the Hilbert space and lead to some applications.

A work connected with Kazdan's property T [42] provides a simple example of continuously many non-isomorphic factors of type II, and answers a question of Murray and von Neumann about the fundamental group of a factor of type II_1.

I hope that I have conveyed to you some feeling about the incredible power of Alain Connes and the richness of his contributions.

References

Published works of Alain Connes

[1] Sur une généralisation de la notion de corps ordonné, *C. R. Acad. Sci. Paris Sér A–B* **269** (1969), pp. A337–A340.

[2] Ordres faibles et localisation des zéros de polynômes, *C. R. Acad. Sci. Paris Sér. A–B* **269** (1969), pp. A373–A376.

[3] Ordres faibles et localisation de zéros de polynômes, *Séminaire Choquet: 1968/69, Initiation à l'Analyse*, Exp. 5, 27 pp. Secrétariat mathématique, Paris, 1969.

[4] Détermination de modèles minimaux en analyse non standard et application, *C. R. Acad. Sci. Paris Sér. A–B* **271** (1970), pp. A969–A971.

[5] Ultrapuissances et applications dans le cadre de l'analyse non standard, *Séminaire Choquet: 1969/70, Initiation à l'Analyse*, Fasc. 1, Exp. 8, 25 pp. Secrétariat mathématique, Paris, 1970.

[6] Un nouvel invariant pour les algèbres de von Neumann, *C. R. Acad. Sci. Paris Sér. A* **273** (1971), pp. 900–903.

[7] Calcul des deux invariants d'Araki–Woods par la théorie de Tomita–Takesaki, *C. R. Acad. Sci. Paris Sér. A* **274** (1972), pp. 175–177.

[8] États presque périodiques sur une algèbre de von Neumann, *C. R. Acad. Sci. Paris Sér. A* **274** (1972), pp. 1402–1405.

[9] Groupe modulaire d'une algèbre de von Neumann, *C. R. Acad. Sci. Paris Sér. A* **274** (1972), pp. 1923–1926.

[10] Une classification des facteurs de type III, *C. R. Acad. Sci. Paris Sér. A* **275** (1972), pp. 523–525.

[11] (Co-author: A. Van Daele) The Group Property of Invariant S of von Neumann Algebras, *Math. Scand.* **32** (1973), pp. 187–192.

[12] Une classification des fracteurs de type III, *Ann. Sci. Ecole Norm. Sup. 4ème Sér.* **6** (1973), pp. 133–252.

[13] Caractérisation des algèbres de von Neumann comme espaces vectoriels ordonnés, *Ann. Inst. Fourier* (Grenoble) **26** (1974), pp. 121–155.

[14] Almost Periodic States and Factors of type III_1, *J. Functional Analysis* **16** (1974), pp. 415–445.
[15] (Co-author: E. J. Woods) Existence de facteurs infinis asymptotiquement abéliens, *C. R. Acad. Sci. Paris Sér. A* **279** (1974), pp. 189–191.
[16] Sur le théorème de Radon–Nikodym pour les poids normaux fidèles semi-finis, *Bull. Sci. Math. 2ème Sér.* **97** (1973), pp. 253–258.
[17] On Hyperfinite Factors of Type III_0 and Krieger's Factors, *J. Functional Analysis* **18** (1975), pp. 318–327.
[18] (Co-author: E. Størmer) Entropy for Automorphisms of Finite von Neumann Algebras, *Acta Math.* **134** (1975), pp. 289–306.
[19] On the Hierarchy of W. Krieger, *Illinois J. Math.* **19** (1975), pp. 428–432.
[20] (Co-author: M. Takesaki) The Flow of Weights on Factors of Type III, *Tōhoku Math. J.* **29** (1977), pp. 473-575.
[21] (Co-author: M. Takesaki) Flot des poids sur les facteurs de type III, *C. R. Acad. Sci. Paris Sér. A* **278** (1974), pp. 945–948.
[22] A Factor not Antiisomorphic to Itself, *Bull. London Math. Soc.* **7** (1975) pp. 171–174.
[23] A Factor not Antiisomorphic to Itself, *Ann. Math.* (2) **101** (1975), pp. 536–554.
[24] Periodic Automorphisms of the Hyperfinite Factor of Type II_1, *Acta Sci. Math.* (Szeged) **39** (1977), pp. 39–66.
[25] The Tomita–Takesaki Theory and Classification of Type III Factors. In: D. Kastler (ed.), *Proceedings of the International School of Physics E. Fermi, Course* **60**, pp. 29–46.
[26] (Co-authors: P. Ghez, R. Lima, D. Testard and E. J. Woods) *Review of Crossed Product of von Neumann Algebras.* Mimeographed notes (unpublished).
[27] Structure Theory for Type III Factors. In: *Proceedings of the International Congress of Mathematicians*, vol. 2 (1975), pp. 87–91.
[28] Classification of Automorphisms of Hyperfinite Factors of Type II_1 and II_∞ and Application to Type III Factors, *Bull. Amer. Math. Soc.* **81** (1975), pp. 1090–1092.
[29] Outer Conjugacy Classes of Automorphisms of Factors, *Ann. Sci. École Norm. Sup. 4ème Sér.* **8** (1975), pp. 383–419.
[30] Sur la classification des facteurs de type II, *C. R. Acad. Sci. Paris Sér. A* **281** (1975), pp. 13–15.
[31] Classification of Injective Factors, *Ann. Math.* **104** (1976), pp. 73–115.
[32] (Co-author: E. Størmer) Homogeneity of the State Space of Factors of Type III_1, *J. Functional Analysis*, **28** (1978), pp. 187–196.
[33] On the Cohomology of Operator Algebras, *J. Functional Analysis* **28** (1978), pp. 248–253.
[34] (Co-author: W. Krieger) Measure Space Automorphisms, the Normalizers of Their Full Groups and Approximate Finiteness, *J. Functional Analysis* **24** (1977), pp. 336–352.
[35] (Co-author: W. Krieger) *Outer Conjugacy of Non-Commutative Bernoulli Shifts.* Mimeographed notes (unpublished).
[36] The Tomita–Takesaki Theory and Classification of Type III Factors. In: D. Kastler (ed.), *C*-algebras and Their Applications to Statistical Mechanics and Quantum Field Theory*, Proc. Int. School of Physics "Enrico Fermi", Varenna, 1973, Ital. Phys. Soc., 1976, pp. 29–46.

[37] The von Neumann Algebra of a Foliation. In: *Mathematical Problems in Theoretical Physics,* Lecture Notes in Phys. 80, Springer Verlag, 1978, pp. 145–151.
[38] Outer Conjugacy of Automorphisms of Factors, *Symposia Math.* Vol. 20, Academic Press, 1976, pp. 149–159.
[39] On the Equivalence Between Injectivity and Semi-Discreteness for Operator Algebra. In: *Colloq. Int. CNRS* no. 274 (Marseille, 1977), 1979, pp. 107–112.
[40] Von Neumann Algebras. In: O. Lehto (ed.), *Proceedings of the International Congress of Mathematicians, Helsinki 1978,* Acad. Sci. Fennica, 1980, pp. 97–109.
[41] (Co-author: E. J. Woods) A Construction of Approximately Finite-Dimensional non-ITPFI factors, *Canad. Math. Bull.* **23** (1980), pp. 227–230.
[42] A factor of Type II_1 with Countable Fundamental Group, *J. Operator Theory* **4** (1980), pp. 151–153.
[43] On the Spatial Theory of von Neumann Algebras, *J. Functional Analysis* **35** (1980), pp. 153–164.
[44] Sur la théorie non commutative de l'integration. In: P. de la Harpe (ed.), *Algèbres d'Operateurs,* Lecture Notes in Math. 725, Springer, 1979, pp. 19–143.
[45] Feuilletages et algèbres d'operateurs, *Séminaire Bourbaki 1979/80,* n° 551, Lecture Notes in Math. 842 (1981), pp. 139–155.
[46] C^*-algèbres et géométrie differentielle, *C. R. Acad. Sc. Paris* **290** A (1980), pp. 599–604.
[47] An Analogue of the Thom Isomorphism for Cross Products of a C^*-algebra by an Action of $\boldsymbol{R}$, *Adv. in Math.* **39** (1981), pp. 31–55.
[48] (Co-author: H. Moscovici) The L^2-Index Theorem for Homogeneous Spaces of Lie Groups, *Ann. of Math.* **115** (1982), pp. 291–330.
[49] (Co-authors: J. Feldmann and B. Weiss) Amenable Equivalence Relations Are Generated by a Single Transformation, *Ergodic Theory Dynamical Systems* **1** (1981), pp. 431–450.
[50] A Survey of Foliations and Operator Algebras, *Proceedings of Symposia in Pure Mathematics* (AMS), **38** (1982), Part 1, pp. 521–628.
[51] (Co-author: G. Skandalis) Theorème de l'indice pour les feuilletages, *C. R. Acad. Sci. Paris, Sér.* I, **292** (1981), pp. 871–876.
[52] (Co-author: V. Jones) A II_1 Factor with Two Nonconjugate Cartan Subalgebras, *Bull. Amer. Math. Soc.* **6** (1982), pp. 211–212.
[53] (Co-author: B. Weiss) Property T and Asymptotically Invariant Subspaces, *Israel J. Math.* **37** (1980), pp. 209–210.
[54] Spectral Sequence and Homology of Currents for Operator Algebras, *Faßungsbericht 42/81,* Oberwolfach (1981), pp. 4–5.
[55] (Co-author: G. Skandalis) *The Longitudinal Index Theorem for Foliations,* IHES Preprint M/82/24 (1982).
[56] Classification des facteurs. In: *Proceedings of Symposia in Pure Mathematics* (AMS), **38** (1982), Part 2, pp. 43–109.
[57] (Co-author: H. Moscovici) L^2-Index Theory on Homogeneous Spaces and Discrete Series Representations, In: *Proceedings of Symposia in Pure Mathematics* (AMS), **38** (1982), Part 1, pp. 419–433.

RESEARCH INSTITUTE FOR MATHEMATICAL SCIENCES
KYOTO UNIVERSITY
KYOTO 606, JAPAN

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

C. T. C. WALL

On the Work of W. Thurston

Thurston has fantastic geometric insight and vision; his ideas have completely revolutionized the study of topology in 2 and 3 dimensions, and brought about a new and fruitful interplay between analysis, topology and geometry.

The central new idea is that a very large class of closed 3-manifolds should carry a hyperbolic structure — be the quotient of hyperbolic space by a discrete group of isometries, or equivalently, carry a metric of constant negative curvature. Although this is a natural analogue of the situation for 2-manifolds, where such a result is given by Riemann's uniformization theorem, it is much less plausible — even counter-intuitive — in the 3-dimensional situation. The case of a manifold fibred over a circle with fibre a surface of genus exceeding 1 seems particularly implausible, and this was the case Thurston examined first. The fibration is determined by a homeomorphism h of the surface, and in seeking to put h (and hence its iterates) into normal form, he was led to consider the images of curves under high iterates of h: these may eventually become dense in some regions, leading to measured foliations. In general, he was led to consider a lamination, which is a disjoint union of injectively immersed curves, which may be dense in some regions and not in others. These ideas gave rise to a geometric model of Teichmüller space and its compactification, which revolutionized thinking in this already highly developed subject.

In this, Thurston was able to draw on his previous work on foliation theory. He swept through this subject producing startling new examples (the Godbillon–Vey invariant takes on uncountably many values), extending the Haefliger foliation theory to closed manifolds by an entirely novel geometric technique, calculating homology of classifying spaces of foliations and relating it to homology of diffeomorphism groups, etc. One dramatic example: any closed manifold of Euler characteristic zero admits a codimension 1 foliation.

The analysis of the diffeomorphisms of a surface concludes with a partition of the surface into well defined pieces, on each of which h has a structure of particularly simple form: the "generic" case is that in which h is Anosov. This partition gives a partition of the total space of the fibration mentioned above, and again we have a geometric structure on each piece. This leads to a reformulation of the project, which occurred at a timely moment in the independent development of 3-dimensional topology.

In the late 1950's, Papakyriakopoulos obtained fundamental new results on embeddings of discs and spheres into 3-manifolds, one consequence of which was a unique decomposition of M by spheres into irreducible pieces. It seemed that these were largely determined by their fundamental groups, and much work went into studying properties of these, though even a basic question like "are the fundamental groups of knots residually finite?" remained unanswered. A method was developed by Haken and Waldhausen using successive decompositions of M by "essential" embedded surfaces to answer such questions: this gave excellent results in the cases to which it applied. These are the manifolds M formerly called "sufficiently large" but now, following Thurston, "Haken 3-manifolds". The condition is that there exists an embedded surface of positive genus whose fundamental group maps injectively to $\pi_1(M)$. If M has a boundary component of positive genus, or if the first Betti number is non-zero, such a surface can be constructed. Yet more particular are the examples given by Seifert fibre spaces. A close analysis of these by Jaco and Shalen and (independently) Johannsen led to a decomposition of M into a "Seifert" piece and an "atoroidal" piece: in the latter, every embedded torus is parallel to the boundary.

Thurston was now able to conjecture that every irreducible, atoroidal 3-manifold has a hyperbolic structure, and to prove it in the case of Haken manifolds. This includes, for example, the complement of any knot in S^3 (other than torus knots and companion knots) — which allows one to prove the residual finiteness mentioned above. It also led to the solution of the Smith Conjecture — that the fixed point set of a periodic homeomorphism of S^3 is always unknotted — a problem which had attracted a great deal of attention over a forty year period. In fact, Thurston formulates the general conjecture in more attractive terms: every compact 3-manifold has a canonical decomposition into pieces each of which has a geometric structure.

Here, a type of geometric structure is defined by a (simply-connected) model manifold X, with a Riemannian metric, and its group G_X of isometries. To say that M has a structure of type X means that there are local coordinate charts from M to X, with transformations between different charts given on their overlaps by elements of G_X. From this, Thurston constructs a developing map φ: $\tilde{M} \to X$ from the universal covering of M: the structure on M is complete if φ is a homeomorphism. In order to be of interest, X must satisfy some conditions (e.g. G_X is transitive): such X he calls *geometries*. Thurston then shows that there are just 8 three-dimensional geometries: in addition to the sphere, Euclidean and hyperbolic space, and products of two-dimensional models with a line, there are 3 further cases, in each of which X is a Lie group. Although the hyperbolic structures are by far the most profound, this general theory of geometric structures has clarified and synthesized much previous work in the other cases also: the list of 8 three-dimensional geometries had not been previously obtained.

The main theorem proved to date is that every compact Haken 3-manifold admits a geometrical decomposition as above. However, hyperbolic structures have also been obtained for numerous non-Haken manifolds. There is also an extended conjecture: that if the manifold has a finite group of automorphisms, then a decomposition, and geometric structures on the pieces, can be found so that the group respects these. This is now known in most cases where a decomposition exists. Of particular difficulty are finite group actions on the 3-sphere. Thurston has shown that most of these are equivalent to orthogonal actions, but the fixed-point free case still eludes the method. Results on these problems have been obtained by other authors using minimal surface theory. Thurston's method involves using his main theorem to obtain a hyperbolic structure on the subset where the group acts freely.

As might be expected, the proof of the general result is long and involves many new ideas: it is not yet all available in detail. All I can do here is to mention a few of the ingredients.

Thurston showed that one can pass from any point in Teichmüller space to any other point by a unique "left earthquake". An example of an earthquake is an incomplete Dehn twist: cut a Riemann surface along a simple closed geodesic and then identify the banks after moving each point on one by the same distance. In a general earthquake the simple closed geodesic is replaced by a lamination. Thurston's earthquake theorem was used by his student Kerckhoff to solve the Nielsen realization problem (every finite subgroup of the Teichmüller modular group has a fixed point).

Discrete isometry groups of hyperbolic 3-space were first studied by Poincaré, who dubbed them "Kleinian groups". These have been much studied by analysts, and Ahlfors' finiteness theorem obtained in the 1960's was a fundamental result. Thurston has studied deformations of such groups (in order to patch together hyperbolic structures defined on two pieces of a manifold): this involves a deep study of limit sets. He has shown that a quasiconformal map on S^2 which conjugates one Kleinian group to another extends to H^3 as a quasiconformal volume preserving map with the same property. Typically, one of these groups is Fuchsian, with limit set S^1, but the image Jordan curve is fractal, with Hausdorff dimensional $d > 1$. It can be constructed as the boundary of a disc obtained by bending the standard D^2 along all the curves of a lamination. Thurston has also shown that for a large class of Kleinian groups (including "degenerate" ones), the limit set has measure zero: thus proving another conjecture which had resisted repeated earlier attempts.

Thurston's work has had an enormous influence on 3-dimensional topology. This area had a strong tradition of "bare hands" techniques, and relatively little interaction with other subjects. Direct arguments remain essential, but 3-dimensional topology has now firmly rejoined the main stream of mathematics.

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

LOUIS NIRENBERG

The Work of Yau, Shing-Tung

Yau has done extremely deep work in global differential geometry and elliptic partial differential equations, including applications in three-dimensional topology and in general relativity theory. He is an analyst's geometer (or geometer's analyst) with remarkable technical power and insight. He has succeeded in solving problems on which progress had been stopped for years. Here are a few of this striking achievements:

1. The Calabi conjecture. Consider a compact Kähler manifold M with Kähler metric $ds^2 = g_{j\bar{k}}dz^j d\bar{z}^k$ and the associated closed form

$$\omega = \frac{i}{2} g_{j\bar{k}} dz^j \wedge d\bar{z}^k.$$

In 1954, Calabi conjectured that given a closed $(1,1)$ form

$$\sigma = \frac{i}{2\pi} \tilde{R}_{j\bar{k}} dz^j \wedge d\bar{z}^k$$

representing the first Chern class of M, there is a Kähler metric $d\tilde{s}^2$ with $\tilde{R}_{j\bar{k}}$ as its Ricci tensor and such that the corresponding $\tilde{\omega} = \frac{i}{2} \tilde{g}_{j\bar{k}} dz^j \wedge d\bar{z}^k$ is in the same cohomology class as ω. Calabi also conjectured the existence (and proved uniqueness) of a canonical Kähler–Einstein metric on a Kähler manifold M with ample canonical line bundle.

Analytically these conjectures reduce to solving a complex Monge–Ampère equation for a real function φ

$$\det\left(g_{j\bar{k}} + \frac{\partial^2 \varphi}{\partial z^j \partial \bar{z}^k}\right) = \det(g_{j\bar{k}}) e^{c\varphi + F}, \quad c = 0 \text{ or } 1,$$

with F a given function, satisfying $\int_M \exp F = \mathrm{Vol}(M)$ in case $c = 0$. In [1] and [3] Yau solved these (and more general) equations by deriving suitable a priori estimates for the solution and its derivatives. The derivation of these estimates, though classical in spirit, is a *tour de force*. Earlier, T. Aubin had proved Calabi's conjecture in case M has non-negative holomorphic bisectional curvature.

Calabi's conjecture has important and beautiful applications in algebraic geometry, some of which are derived in [1]. For example:

(i) The only Kähler structure on a complex projective space is the standard one.

(ii) There are compact simply connected Kähler manifolds whose Ricci curvature is everywhere negative, for example, any complex hypersurface of CP^{n+1} with degree $> n+2$.

(iii) Suppose M is a Kähler surface with ample canonical bundle. Then $3C_2(M) \geqslant C_1^2(M)$, and equality holds iff M is covered holomorphically by the ball in C^2.

(iv) (Conjectured by Severi) If a complex surface is homotopic to CP^2, then it is biholomorphic to CP^2.

2. The positive mass conjecture. This asserts that for a nontrivial isolated physical system (in general relativity theory), the total energy, including contributions from matter and gravitation, is positive. This was first proved by R. Schoen and Yau [12, 13]; a simpler proof was later given by E. Witten. Their proof involves the construction of global minimal surfaces and a study of their stability and behaviour near infinity. It is very technical and extremely ingenious. Recently in [14] they adapted their arguments to prove the positivity of the Bondi mass — the total mass of the isolated physical system measured after loss by gravitational radiation.

3. Real and complex Monge-Ampère equations. In joint work with S. Y. Cheng (based partly on work of A. V. Pogorelov), Yau gave a complete proof of the higher dimensional Minkowski problem (to determine a closed convex hypersurface in R^{n+1} if its Gauss curvature is given as a function of its normal), as well as the Dirichlet problem for the real Monge–Ampère equation [7], [8]. In [9] they carry over some of the results of [3] to noncompact complex manifolds. They prove that if a complex manifold with a complete Kähler metric satisfies: (i) its Ricci curvature is $\leqslant C_0 < 0$, (ii) its injectivity radius is $\geqslant C_1 > 0$, (iii) its curvature tensor and its covariant derivatives are bounded, then the manifold admits a unique Einstein–Kähler metric. Furthermore, in a strictly pseudo-convex domain in C^n with smooth boundary, they study the behaviour near the boundary of the unique Einstein–Kähler metric. This involves the study of the solution of the Fefferman equation

$$(-1)^n \det \begin{bmatrix} v & v_{\bar{k}} \\ v_j & v_{j\bar{k}} \end{bmatrix} = 1$$

with $v = 0$ on the boundary. This work involves deep, technical estimates.

4. Various attempts have been made to find higher dimensional forms of the uniformization theorem. In this connection there was the Frankel conjecture that a complete, simply connected Kähler manifold M, with positive holomorphic bisectional curvature, is biholomorphic to a complex projective space. This was proved by S. Mori using methods from algebraic topology of characteristic p.

Y. T. Siu and Yau [16] gave a very elegant analytic proof with the aid of harmonic maps. Earlier [15] they had proved that if, on the other hand, the sectional curvatures of M satisfy

$$0 \geqslant \text{sectional curvature} \geqslant -\frac{A}{1+r^{2+\varepsilon}},$$

A, $\varepsilon > 0$, and r = distance from some point of M, then M is biholomorphic to C^n.

5. Further work on elliptic equations and geometry. In a series of papers, some with P. Li (see [10], where other references may be found) Yau obtained estimates on the first and other eigenvalues for the Laplace operator on a compact manifold, with or without boundary, under various hypotheses on the Ricci curvature — but relying on little further information about the Riemannian manifold.

Yau proved a very useful form of the maximum principle in non-compact manifolds M and used it to derive various geometric results. In his paper [6] he proved there is no non-constant positive harmonic function on a complete Riemannian manifold with non-negative Ricci curvature.

6. Connections with topology. In a series of papers with W. H. Meeks III (see [11] which contains other references) Yau used topological methods of 3-manifolds to settle some old problems for minimal surfaces. Conversely, they used minimal surface theory to derive results in 3-dimensional topology, such as Dehn's lemma and equivariant versions of the loop and sphere theorems. Among a certain family of maps of the disc, or sphere, into M^3, they show there is one with minimal area. Then using the tower construction in topology, they prove that any area minimizing map in the family is an embedding — thus realizing solutions of topological problems as minimal surfaces.

The equivariant loop theorem together with a theorem of Thurston were used to prove the Smith conjecture.

In [5] with H. Blaine Lawson, Yau proved that if a compact manifold admits a smooth action by a compact, connected, non-abelian Lie group, then it admits a metric of positive scalar curvature. They then prove that if Σ^n is an exotic n-sphere which does not bound a spin manifold, then the only possible compact connected transformation groups of Σ^n are tori of dimension $\leqslant [(n+2)/2]$.

Paper [2] is an interesting study of the group of isometries of a Riemannian manifold.

Yau has many papers concerned with minimal surfaces; he uses these surfaces in the way that, previously, people had used geodesics.

Finally, for a good picture of the role of analysis in differential geometry, we recommend the two articles by Yau in the seminar [4], the first, which is a survey, and the last, a list of open problems in differential geometry — many contributed by Yau himself.

Yau's work covers all aspects of global differential geometry and is distinguished by great technical power, and depth, as well as by its wide variety — also by his courage and vision.

References

Works of S. T. Yau

[1] Calabi's Conjecture and Some New Results in Algebraic Geometry, *Proc. Nat. Acad. Sci. USA* **74** (1977), pp. 1798–1799.

[2] Remarks on the Group of Isometries of a Riemannian Manifold, *Topology* **16** (1977), pp. 239–247.

[3] On the Ricci Curvature of a Compact Kähler Manifold and the Complex Monge–Ampère Equation I, *Comm. Pure Appl. Math.* **31** (1978), pp. 339-411.

[4] *Seminar on Differential Geometry,* edited by S. T. Yau, Annals of Math. St. No. 102, Princeton Univ. Press, Princeton, N. J., 1982.

[5] with H. Blaine Lawson, Jr., Scalar Curvature, Non-Abelian Group Actions, and the Degree of Symmetry of Exotic Spheres, *Comment. Math. Helv.* **49** (1974), pp. 232–244.

[6] Harmonic Functions on Complete Riemannian Manifolds, *Comm. Pure Appl. Math.* **28** (1975), pp. 201–228.

[7] with S. Y. Cheng, On the Regularity of the Solution of the n-Dimensional Minkowski Problem, *Comm. Pure Appl. Math.* **29** (1976), pp. 495–515.

[8] with S. Y. Cheng, On the Regularity of the Monge–Ampère Equation det $(\partial^2 u/\partial x_i \partial x_j) = F(x, u)$, *ibid.* **30** (1977), pp. 41-68.

[9] with S. Y. Cheng, On the Existence of a Complete Kähler Metric on Non-Compact Complex Manifolds, and the Regularity of Fefferman's Equation, *ibid.* **33** (1980), pp. 507–544.

[10] with P. Li, Estimates of Eigenvalues of a Compact Riemannian Manifold, *Proc. Symp. Pure Math.* **36** (1980), pp. 205-239.

[11] with W. H. Meeks III, The Existence of Embedded Minimal Surfaces and the Problem of Uniqueness, *Math. Z.* **179** (1982), pp. 151-168.

[12] with R. Schoen, On the Proof of the Positive Mass Conjecture in General Relativity, *Comm. Math. Phys.* **65** (1979), pp. 45-76.

[13] with R. Schoen, On the Proof of the Positive Mass Conjecture II, *Comm. Math. Phys.* **79** (1981), pp. 231-260.

[14] with R. Schoen, *The Proof of the Positivity of the Bondi Mass.*

[15] with Y. T. Siu, Complete Kähler Manifolds with Nonpositive Curvature of Faster than Quadratic Decay, *Ann. of Math.* **105** (1977), pp. 225–264.

[16] with Y. T. Siu, Compact Kähler Manifolds of Positive Bisectional Curvature, *Invent. Math.* **59** (1980), pp. 189-204.

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

JACOB SCHWARTZ

The Work of Robert Endre Tarjan

Pure mathematics enjoys the luxury of studying its constructions, whether finite or infinite, in complete independence of all questions of efficiency. By contrast, theoretical computer science must ultimately concern itself with computing engines which operate with limited speed and data storage, and therefore must take efficiency as one of its central concerns. Two closely related activities, algorithm design and algorithm analysis, grow out of this inevitable concern with efficiency. The *algorithm designer* aims to find procedures which accomplish specified computational tasks as efficiently as possible; *algorithm analysis* attempts to establish upper bounds on attainable performance, which, as they become realistic, provide optimality targets at which algorithm designers can aim.

Robert Tarjan has been a leader in both these enterprises, which lie at the intellectual heart of computer science, and which during the past two decades have begun to define the enduring theoretical inheritance which computer science will be able to carry forward. He is perhaps the strongest designer of high-efficiency combinatorial algorithms active today, and deserves much of the credit for raising the field to its present level of sophistication. He has succeeded in devising near-optimal algorithms for many graph-theoretic and geometric problems which had at first appeared quite intractable, and has also contributed several algorithmic analyses of striking profundity and elegance. His work is distinguished by a meticuluos craftsmanship, which returns repeatedly to crucial problems, always to replace refined results by constructions more nearly perfect.

Several themes have characterized his work. One important thread is that of effective choice of processing order for structures, such as graphs, whose basic definition suggests no particular ordering. Here he has devised, and many times exploited, the simple but remarkably effective concept of depth-first spanning tree for a graph. This is any subgraph produced by stepping from node to adjacent unvisited node whenever possible, but retreating to nodes visited earlier when nodes having no unvisited neighbours are encountered. The ordering of nodes that results from this simple graph-traversal procedure turns out to provide a remarkably advantageous starting point for other graph analysis problems.

A pioneering demonstration of the efficiency attainable by use of this technique appears in Tarjan's first but remarkably mature work (with J. Hopcroft) on the classical problem of graph planarity testing. The same depth-first tree theme, which has become a kind of Tarjan trademark, was continued, but used in new and striking ways, in subsequent papers on the problem of graph dominators and on graph reducibility, which has significantly influenced some aspects of practical compiler-writing.

Tarjan has also applied other, quite different, graph orderings which proved effective for other graph-theoretic problems, some related to major problems in numerical analysis, e.g. the choice of optimal elimination orders for solution of sparse systems of linear equations.

A second major theme of Tarjan's work has been development and exploitation of data structures capable of supporting various special combinations of elementary operations with remarkable efficiency. Data structures of this kind play an essential role in many of his most refined algorithms. In this connection he has developed and exploited new forms of rapidly modifiable search trees which improve access speed by exploiting information concerning the frequency with which particular items will be accessed; new techniques for maintaining collections of trees as they are dynamically joined and cut into subtrees; efficient representations of priority queues, etc. The design of these data structures has gone hand in hand with development of clever counting techniques for making their efficacy manifest.

To pave the way for application of these ingenious graph-processing methods and remarkable data structures, Tarjan has repeatedly found deep and surprising transformations of graph-related problems. This additional level of combinatorial orchestration is seen in his work (with Hopcroft) on efficient analysis of graphs into triconnected components, on decompositions of planar graphs in ways facilitating application of finite element techniques, and generally in a remarkably prolific outpouring of work on graph-theoretic and geometric algorithms.

Yet another theme of Tarjan's work is the application of sophisticated enumerative combinatorics to analyse the efficiency of particular algorithms, and, more profoundly, to estimate the maximum efficiency which can be attained by whole classes of algorithms which he has studied. In this area his work (with his student T. Lengauer and others) on the "pebbling game" model of computation represents one of the most sophisticated and successful elucidations of the deep problem of time/space tradeoffs in computation.

The mass of his work shows his enthusiasm, energy, and force; the exceptionally many major combinatorialists and computer scientists with whom he has collaborated show the warmth and generosity of his scientific spirit. He sets a high standard of excellence for the new prize now awarded to him; if future recipients of this prize meet the same standard, computer science will do well indeed.

Invited One-Hour Plenary Addresses

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

V. I. ARNOLD

Singularities of Ray Systems

The simplest example of a ray system is the system of all normals to a given surface in Euclidean space. Hamilton (1824) turned the theory of ray system into a part of symplectic geometry; since Maslov's thesis (1965) ray systems are called Lagrangian submanifolds.

The normals to a surface foliate some neighbourhood of that surface; but away from that neighbourhood various normals start intersecting one another (Fig. 1). The resulting complicated and beautiful geometry

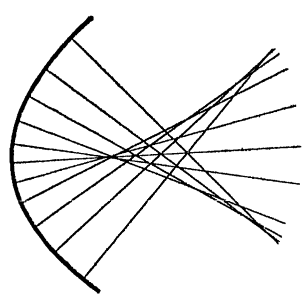


Fig. 1

was hidden up to 1972, when the relation between singularities of ray systems and Euclidean reflection groups was discovered.

This relation, for which there is no a priori reason, turned out to be a powerful method for the analysis of singularities. By 1978 it became clear that the Euclidean reflection groups also govern the singularities of Huygens evolvents.

Huygens (1654) discovered that the evolvent of a plane curve has a cusp singularity at each point of contact with the curve (Fig. 2). Plane curve evolvents and their higher-dimensional generalizations are the wave fronts on manifolds with boundary. The singularities of wave fronts, as well as those of ray systems, are classified by reflection groups.

While the ray and front systems on manifolds without boundary are related to the A, D and E series of the Weyl groups, the singularities of evolvents are described by the B, C, F series (those having Dynkin diagrams with double connections).

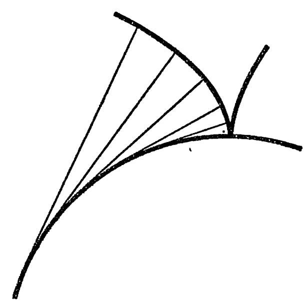


Fig. 2

The relation of the remaining reflection groups $(I_2(p), H_3, H_4)$ to singularity theory was unknown until recently. This situation has changed since the fall of 1982 when it was discovered that the group H_3 (the group of symmetries of the icosahedron) governs the singularities of evolvent systems at the inflection points of plane curves.

The icosahedron appears at an inflection point as mystically as it does in Kepler's law of planetary distances. I believe, however, that in our problem the appearance of the icosahedron is more relevant than in Kepler's case; I hope that the remaining group H_4 will appear naturally in the analysis of the more complicated singularities of ray systems and wave fronts.

The main theme of this paper is the application of the relation between singularities of ray systems and reflection groups. The results I shall discuss are now included in symplectic and contact geometries under the names of Lagrangian and Legendrian singularity theories. But one may consider them as part of the calculus of variations, or of control theory, of PDE theory, or of classical mechanics, of optics, or of wave theory, of algebraic geometry, or of general singularity theory. Some of these results deal with objects so basic, that it seems strange that the classics have missed them. For example, the local classification of projections of surfaces in general position in the usual 3-space was discovered only in 1981.

The number of nonequivalent projection germs is 14: the point neighbourhoods on generic surfaces generate exactly 14 different patterns when the surfaces are seen from different points of 3-space.

The reason is perhaps the difficulty of the proofs: they depend on the relations (sometimes unexpected) to invariant theory, Lie algebras, reflexion groups, algebraic geometry, and Deligne mixed Hodge structures. Some of the results were stimulated by applications of singularity theory to perturbation analysis of Hamiltonian dynamical systems, and even to number theory, but most new concepts came from the problem of bypassing an obstacle in Euclidean 3-space.

In order to describe these new results I must recall some well-known notions.

1. Symplectic geometry

A *symplectic structure* on an even-dimensional smooth manifold is a closed nondegenerate differential 2-form on it.

Examples: 1. The oriented area element defines a symplectic structure on the plane. 2. The direct product of symplectic manifolds has a natural symplectic structure. 3. The phase space of classical mechanics (the total space of the cotangent bundle of a smooth manifold) has a natural "$dp\wedge dq$" symplectic structure. 4. One may equip the manifold of oriented lines in Euclidean space with the symplectic structure of the total space of the cotangent bundle of the sphere, since these two manifolds are diffeomorphic. 5. The *characteristic direction* at a point of a hypersurface in a symplectic manifold is the skew-ortho-complement to the tangent plane. The *characteristics* on a hypersurface are the integral lines of its field of characteristic directions. The manifold of characteristics inherits a symplectic structure from the original manifold. 6. In particular, the manifold of extremals of general variational problem, lying at the same level manifold of the Hamiltonian function, is equipped with a natural symplectic structure. 7. Consider the space of odd-degree binary forms. There exists a unique (up to constant multiple) nondegenerate SL_2-invariant bilinear skew form on this even-dimensional linear space. This form defines a natural symplectic structure on the space of binary forms. 8. The binary forms in x and y, with coefficient in front of x^{2k+1} equal to 1, form a hyperplane in the space of all forms. The manifold of characteristics of this hyperplane can be identified with the manifold of even-degree polynomials in x of the form $x^{2k}+\ldots$ We have thus equipped this space of even-degree polynomials with a symplectic structure. 9. The one-parameter group of shifts along the x-axis preserves this symplectic structure. The Hamiltonian function of this group is a polynomial of degree 2, known already to Hilbert (1893). The manifold of characteristics of a level hypersurface of the Hamiltonian function can be identified with the manifold of polynomials $x^{2k-1}+\dots$ with sum of roots equal to 0. We thus get a natural symplectic structure on this space of polynomials.

THEOREM (G. Darboux, 1882). *All the symplectic structures on manifolds of a fixed dimension are locally diffeomorphic.*

Thus, every symplectic structure is locally reducible to the normal form $\sum dp_i \wedge dq_i$ by a suitable choice of local "*Darboux coordinates*" p_i, q_i.

Let us now consider submanifolds of a symplectic manifold. The restriction to the submanifold of the symplectic structure is a closed 2-form, but it is not necessarily nondegenerate. In Euclidean space there is not only the inner geometry of a submanifold, but also an extensive theory of exterior curvatures. In the symplectic case the situation is much simpler:

THEOREM (A. B. Givental, 1981). *The germ of a submanifold of a symplectic manifold is determined (up to a symplectic diffeomorphism) by the restriction of the symplectic form to the tangent spaces of the submanifold.*

An intermediate theorem, dealing with vectors nontangent to the submanifold, was proved by A. Weinstein (1973). Unlike the Weinstein theorem, the Givental theorem implies the classification of the germs of generic submanifolds in a symplectic space: one uses the classification of the degeneracies of symplectic structures obtained by J. Martinet (1970) and his followers.

Examples: 1. The germs of a generic 2-surface in a symplectic manifold are locally symplectomorphic (symplectically diffeomorphic) to those of the surface $p_2 = p_1^2$, $q_1 = 0$, $p_3 = q_3 = \dots = 0$ (we use the Darboux coordinates). 2. On a 4-submanifold one encounters stably the curves of elliptic and hyperbolic Martinet singularities with normal forms

$$p_2 = p_1p_3 \pm q_1q_2 + q_3^3/6, \quad p_3 = 0, \quad p_4 = q_4 = \dots = 0.$$

The ellipticity and hyperbolicity concern the character of the motions in a dynamical system related intrinsically to the submanifold. The relevant divergence-free vector field on a 3-dimensional manifold has a curve of singular points. The classification at singular curves turns out to be less pathological than that at singular points (the latter being almost as difficult as the whole of celestial mechanics).

I have thus described the first steps of the symplectic singularity theory of smooth submanifolds.

A *Lagrangian submanifold* of a symplectic manifold is a submanifold on which the restriction of the symplectic structure vanishes, and which has highest possible dimension (equal to half of the dimension of the symplectic manifold).

Examples: 1. The fibers of the cotangent bundle. 2. The manifold of lines normal to a smooth submanifold (of arbitrary dimension) in Euclidean space. 3. The set of all polynomials $x^{2m}+\dots$ divisible by x^m.

A *Lagrangian fibration* is a fibration whose fibers are Lagrangian submanifolds.

Examples: 1. The cotangent bundle. 2. The fibration sending an oriented line in Euclidean space to the corresponding unit vector at the origin.

All Lagrangian fibrations of a given dimension are locally symplectomorphic (in the neighbourhood of each point of the total space).

A *Lagrangian mapping* is a diagram $V \to E \to B$, where the first arrow is an immersion of a Lagrangian submanifold, and the second is a Lagrangian fibration (Fig. 3).

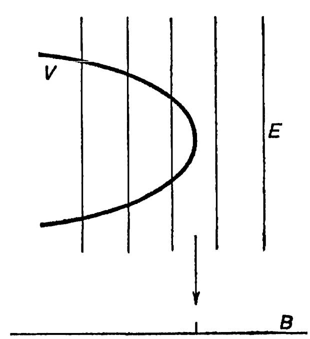


Fig. 3

Examples: 1. The gradient mapping: $q \mapsto \partial S/\partial q$. 2. The normal mapping: associate to each vector normal to a submanifold in Euclidean space its end point. 3. The Gaussian mapping: associate to each point of a transversally oriented hypersurface in Euclidean space the unit vector at the origin in the direction of the normal at that point. (The corresponding Lagrangian manifold consists of the normals to that hypersurface.)

An *equivalence* between Lagrangian mappings is a fiber-preserving symplectomorphism between the total spaces of the fibrations, mapping the first Lagrangian submanifold onto the second one.

The set of critical values of a Lagrangian mapping is called its *caustic*. The caustics of equivalent mappings are diffeomorphic.

Example. The caustic of the normal mapping of a surface is the envelope of its normals, i.e., its focal surface (the surface of the curvature centers).

Every Lagrangian mapping is locally equivalent to a gradient one (to a normal one, to a Gaussian one). The singularities of generic gradient (normal, Gaussian) mappings are equivalent to those of generic Lagrangian mappings. These singularities are classified by the Euclidean reflection groups A, D, E.

Example. Consider a medium of dust-like particles moving inertially whose velocities form a potential field. After a time interval t a particle moves from x to $x+t\,\partial S/\partial x$. We obtain a one-parameter family of smooth mappings $\boldsymbol{R}^3 \to \boldsymbol{R}^3$.

These are Lagrangian mappings. Indeed, a potential field of velocities defines a Lagrangian section of the cotangent bundle. The phase flow of Newton's equation sends the initial Lagrangian manifold to new Lagrangian manifolds, which, however, need not be sections (for large t): their projections to the base space may have singularities (Fig. 4). The caustics

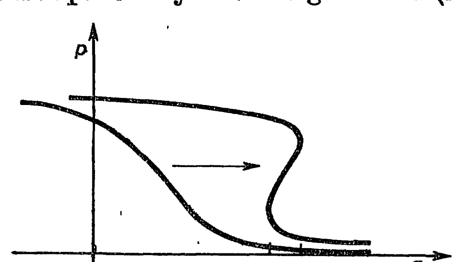


Fig. 4

of these mappings are the places where the density of particles becomes infinite. According to Ya. B. Zel'dovich (1970), a similar model (taking into account gravitation and expansion of the Universe) describes the generation of the large-scale nonuniformity of the distribution of matter in the Universe.

The theory of Lagrangian singularities implies that a new-born caustic has the shape of a saucer (at moment t after its birth the saucer's axes

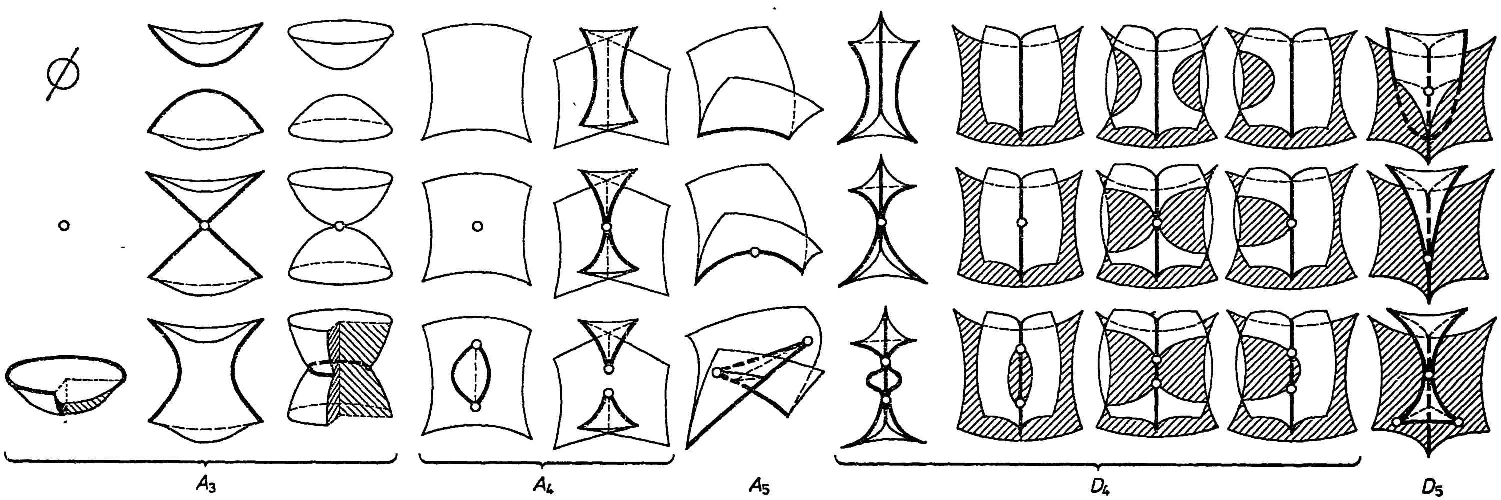


Fig. 5

are of order $t^{1/2}$, its depth of order t and thickness of order $t^{3/2}$). The saucer's birth corresponds to A_3. All metamorphoses of caustics in generic one-parameter families of Lagrangian mappings in 3-space are presented in Fig. 5 (1976).

THEOREM (1972). *The germs of generic Lagrangian mappings of manifolds of dimension $\leqslant 5$ are stable and simple (have no moduli) at every point. The simple stable germs of Lagrangian mappings are classified by the A, D, E Euclidean reflection groups, as explained below.*

2. Contact geometry

A *contact structure* on an odd-dimensional smooth manifold is a nondegenerate field of hyperplanes in the tangent spaces. The exact meaning of "nondegenerate" is irrelevant because of the "Darboux contact theorem": in the neighbourhood of a generic point, all generic fields of hyperplanes on a manifold of a fixed odd dimension are diffeomorphic.

Examples: 1. The *space of contact elements* of a smooth manifold consists of all its tangent hyperplanes. The velocity of an element belongs to the hyperplane defining the contact structure, if and only if the velocity of the contact point belongs to that element. 2. The space of 1-jets of functions $y = f(x)$ has a natural contact structure $dy = p\,dx$ ($p = \partial f/\partial x$ for the 1-jet of $y = f(x)$ at x).

The external geometry of a submanifold of a contact space is locally determined by the internal one, i.e., by the contact structure traces on the tangent spaces (the Givental contact theorem).

An integral submanifold of a contact manifold is said to be *Legendrian* if it has the highest possible dimension.

Examples: 1. The set of all contact elements tangent to a fixed submanifold (of arbitrary dimension). 2. In particular, the contact elements at a given point form a Legendrian manifold (the fibre of the bundle of contact elements). 3. The set of 1-jets of a function.

A fibration is said to be *Legendrian* if its fibers are Legendrian submanifolds.

Examples: 1. The projective cotangent fibration (a contact element is sent to its contact point). 2. The fibration of 1-jets of a function over its 0-jets (forgetting derivatives).

All Legendrian fibrations of a given dimension are locally contactomorphic (at every point of the total space).

The projection of a Legendrian submanifold to the base of a Legendrian fibration is called a *Legendrian mapping.* Its image is called a *front.*

Examples: 1. *The Legendre transformation*: A hypersurface in a projective space can be lifted to the space of its contact elements as a Legendrian submanifold. The manifold of contact elements of the projective space fibers over the dual projective space (associate to a contact element the hyperplane containing it). This fibration is Legendrian. The projection maps the lifted Legendrian manifold to the hypersurface which is projectively dual to the original hypersurface. Thus the projective dual of a smooth hypersurface is a Legendrian mapping front. 2. *The equidistant mapping*: Pick a point on every oriented normal to a hypersurface in Euclidean space, at distance t from the hypersurface (along the normal). We get a Legendrian mapping whose front is equidistant from the given hypersurface.

Legendrian equivalence, stability and simplicity are defined by analogy with the Lagrangian case.

Every Legendrian mapping is locally equivalent to a mapping defined by a Legendre transformation, and to an equidistant mapping. The local Legendrian singularity theory coincides with that of singularities of Legendre transformations (or equidistant mappings, or wave fronts).

THEOREM (1973). *The germs of generic Legendrian mappings of manifolds of dimension $\leqslant 5$ are stable and simple at every point. The simple stable germs of Legendrian mappings are classified by the A, D, E Euclidean reflection groups: the Legendrian mapping fronts are holomorphically equivalent to the varieties of singular orbits of the corresponding reflection groups.*

Example. The singularities of a generic wave-front in 3-space are (semicubical) cuspidal edges (A_2), and swallow-tails (A_3, Fig. 6: at these points the front is diffeomorphic to the surface in the space of polynomials x^4+ax^2+bx+c, consisting of the polynomials having multiple roots).

Remark. The necessity to complexify in the above theorem suggests that Euclidean reflection groups may have different real forms.

All Lagrangian singularities can be constructed from the Legendrian ones. For this, one considers Legendrian submanifolds of the space of 1-jets of functions. By forgetting the value of the functions one projects the jet space onto the phase space. The Legendrian manifolds' germs are projected isomorphically onto the Lagrangian ones. For instance, the caustic of a Lagrangian mapping is the projection of the cuspidal edge of the Legendre mapping front under a generic projection with 1-dimensional fibers.

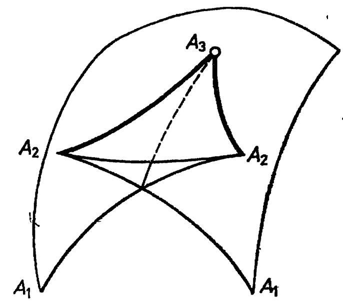


Fig. 6

THEOREM (O. V. Lyashko, 1979). *All holomorphic vector fields transversal to a front of a simple singularity can be mapped one onto another by front-preserving holomorphic diffeomorphisms germs.*

Example. A generic vector field in a neighbourhood of the "most singular point" of the swallow-tail $\{x^4+ax^2+bx+c = (x+a)^2 \ldots\}$ is reducible to the normal form $\partial/\partial c$ (Fig. 7) by a swallow-tail-preserving diffeomorphism.

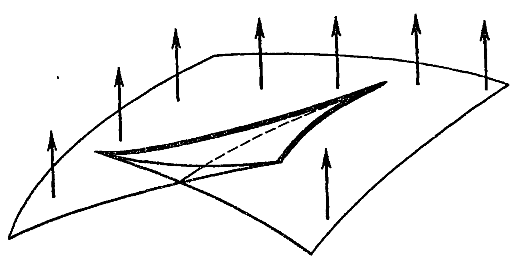


Fig. 7

The reduction to normal form of various geometric objects by wave-front or caustic-preserving diffeomorphisms is the main technical tool in the geometry of ray systems and wave fronts. For instance, the study of the metamorphoses of moving wave fronts is based on a result "dual" to the Lyashko theorem.

THEOREM (1976). *Generic holomorphic functions equal to 0 at the "most singular" point of a simple singularity front can be mapped one onto another by front-preserving holomorphic diffeomorphisms germs.*

Example. A generic function at the most singular point of a swallow-tail is reducible to the normal form a+const by a swallow-tail-preserving diffeomorphism.

The theorem above follows from the equivariant Morse lemma. We use it as follows. The momentary wave fronts form a "large front" in the space-time. Reduce the time function in the space-time to normal form by a large-front-preserving diffeomorphism. We obtain the normal form of the metamorphosis of the momentary front.

The infinitesimal diffeomorphisms preserving a front are the vector fields tangent to it. Their study leads to a "convolution operation" on the invariants of the reflection group. This operation associates to a pair of invariants (i.e., of functions on the orbit space) a new invariant — the scalar product of the gradients of the given functions (lifted from the orbit space to the Euclidean space).

The linearization of this operation is a bilinear symmetric operation on the space cotangent to the orbit space at 0.

THEOREM (1979). *The linearized convolution of the invariants is equivalent to the operation $(p, q) \mapsto S(p \cdot q)$ on the local algebra of the corresponding singularity, where $S = D + (2/h)E$, h is the Coxeter number, and D is the Euler quasihomogeneous derivation.*

For the exceptional groups this theorem was proved by A. B. Givental. In his joint work with A. N. Varchenko (1981) the theorem is extended to higher quasihomogeneous singularities. In this extension they substitute the Euclidean structure by the intersection form of a suitable nondegenerate period mapping. This period mapping comes from a family of holomorphic differential forms on the fibers of the Milnor fibration associated to a versal deformation of a function. A nondegenerate intersection form determines (according to the parity of the number of variables of the function) either a locally flat pseudo-Euclidean metric with a standard singularity at the Legendrian front, or a symplectic structure which is holomorphically extendable to the front.

Example. The set of odd-degree polynomials having highest coefficient equal to 1 and sum of the roots equal to 0 is thus equipped with a new symplectic structure. The variety of polynomials with maximal possible number of double roots is a Lagrangian subvariety.

3. Applications of Lagrangian and Legendrian singularities

The theory was first developed for the study of asymptotics of oscillatory integrals by the stationary phase method. I shall not discuss these (very important) applications here in detail, but shall rather mention: (1) Varchenko's (1976) proof of the formula describing the exponent of the main term of oscillatory integrals in terms of the Newton boundary of the phase function; (2) the example due to the same author of the nonsemicontinuity of this exponent; and (3) V. N. Karpushkin's (1981) proof of a uniform with respect to the parameters estimate from above of the double oscillatory integrals (for simple integrals such an estimate was obtained by I. M. Vinogradov, and for triple ones it was disproved by Varchenko's nonsemicontinuity example).

The uniform estimate also holds for all members of generic families of functions depending on a small number l of parameters (Duistermaat proved it in 1974 for $l \leqslant 6$; Colin de la Verdiere in 1977 for $l \leqslant 7$; Karpushkin in 1982 for $l \leqslant 9$); $l = 73$ is too large (the Varchenko example becomes possible).

The study of asymptotic expansions of oscillatory integrals in the complex domain has led Varchenko (1980–1981) to the construction of a mixed Hodge structure, which he calls the *asymptotic structure*. He has proved that its Hodge numbers coincide with the mixed Hodge numbers constructed algebraically by Steenbrink (1976). Among the corollaries of Varchenko's theory are: (1) the constancy of the Hodge structure invariants along the "$\mu = \text{const}$" statum, and (2) the fact that the "inner modality" of quasihomogeneous functions coincides with their true modality. In real algebraic geometry the mixed structure gives some generalizations of the Petrovskii–Oleinik inequalities.

THEOREM (1978). *The local Poincaré index of a gradient vector field in* $\boldsymbol{R}^{2n}$ *is bounded from above by the middle Hodge number* $|\text{ind}| \leqslant h_1^{n,n}$.

The singularity mixed Hodge structure associates to a finite multiplicity critical point of a function a finite set of rational numbers, the critical points *spectrum*. The spectrum's left end is the smallest exponent of the oscillatory integrals with a given phase function (along complex chains). The examples show the semicontinuity of this exponent, as well as of all the other spectrum points. For instance, the spectrum obtained through a deformation reducing the multiplicity by one, divides the initial spectrum (in the same way as the axes of an ellipsoid divide the axes of the initial ellipsoid).

The spectrum semicontinuity conjecture (1978) was recently confirmed by the works on an apparently unrelated to it algebraic geometry problem: *how large can the numbers of (Morse) singular points on a hypersurface of degree d in CP^n be?*

Bruce (1981) gives an estimate from above, which asymptotic (for the surfaces in CP^3) is $d^3/2 + \ldots$ (the best estimates from below are of order $3d^3/8$, S. Chmutov, 1983). Comparing the first exactly known answers (0, 1, 4, 16, 31, 64) with the mixed Hodge structure, I have formulated the following

CONJECTURE. *The number of singular points does not exceed the number of integer points m of the cube $(0, d)^n$, for which $(n-2)d/2+1 < \sum m_i \leqslant nd/2$.*

For surfaces in 3-space this implies an estimate from above $23d^3/48 + \ldots$ Trying to prove this cojecture A. B. Givental in October of 1982 improved the lower order terms in the Bruce estimate. His proof uses some Rayleigh–Fisher–Courant type inequalities and makes transparent the relation of the problem to the spectrum semicontinuity conjecture.

A. N. Varchenko immediately applied to this problem the Steenbrink (1976) theorem on the limits of the Hodge structures. Thus he proved both the conjectured estimate of the number of singular points and the spectrum semicontinuity (the last — for quasihomogenous function deformations, generated by adding lower weight monomials). The same way he proved the semicontinuity of the left end of the spectrum for all functions in 3 variables and for functions in n variables having "far away" Newton polyhedra.

I shall also mention the applications of Lagrangian singularities to the mechanical quadrature theory, i.e., to the problem of integer points in large domains. Let V be the volume of a smooth boundary domain G in the Euclidean $\boldsymbol{R}^n$, and $N(\lambda)$ the number of integer points inside λG, $R(\lambda) = \lambda^n V - N(\lambda)$. The Lagrangian singularity theory implies the following results:

THEOREM (Colin de Verdiere, 1977). *For $n \leqslant 7$ generically*

$$|R(\lambda)| \leqslant C\lambda^{n-2+2/(n+1)}.$$

THEOREM (Varchenko, 1981). *The average $|R(\lambda)|^2$ over all lattices obtained from the integer point lattice by rotations and shifts, does not exceed $C\lambda^{n-1}$.*

The convex analytic case was studied by Randol (1969). The exponent $(n-1)/2$ is what one might expect according to the law of large numbers (if the λ^{n-1} cells were divided by the boundary independently). The proof of the last theorem is inspired by the Duistermaat (1974) proof of the Maslov "canonical operator" unitarity.

The statistics of Newton diagrams of singularities has led to another inequality related to integer points.

THEOREM (K. A. Sevastianov, S. V. Konyagin, 1982). *The number of vertices of a volume V convex polyhedron in $\boldsymbol{R}^n$, whose vertices are integer points, does not exceed $CV^{(n-1)/(n+1)}$ (the same estimate holds also for the number of faces of arbitrary dimension).*

The influence of the boundary inflections on the remainder term of the asymptotic of the number of integer points is a particular case of the interrelations between the integer and smooth structures of $\boldsymbol{R}^n$, which are crucial for many branches of calculus.

For instance, the order of approximation of a typical point of a submanifold by the hyperplanes defined by the equations with not too large integer coefficients is essential for the resonance phenomena in the theory of nonlinear oscillations (the flattening of the fast frequencies' manifold enhances the sticking at resonances).

In his study of evolutions of action variables in Hamiltonian systems, N. N. Nehoroshev introduced "steepness exponents" of the unperturbed Hamiltonian function. The calculation of these exponents for a generic Hamiltonian function has inspired the theory of tangential singularities.

4. Tangential singularities

These are singularities of the arrangement of a projective surface with respect to its tangents of all dimensions.

Example. The tangential classification of points on a generic surface in 3-space (Fig. 8) was found by O. A. Platonova and E. E. Landis (1979). A line (p) of parabolic points divides the surface into the domain (e) of elliptic points and that of hyperbolic ones (h) containing the curve (f) of inflection points of the asymptotic lines with the biinflection points (b), the selfintersection points (c), and the points of tangency to the parabolic line (t).

This classification is useful both for Nehoroshev's exponent estimate and for the classification of projection degenerations.

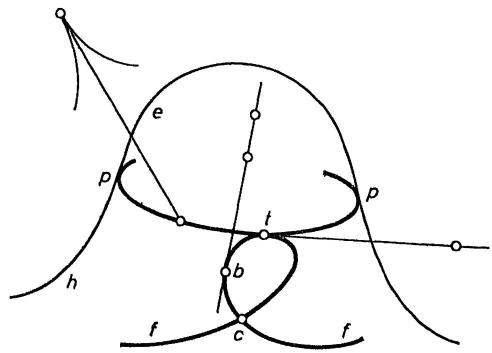


Fig. 8

THEOREM (O. A. Platonova, O. P. Shcherbak, 1981). *Project a generic surface from* $\boldsymbol{RP}^3$ *to a plane along the straight lines passing through a projection center (a point outside the surface).*

All the projections thus obtained are locally equivalent to the 14 *projections of surfaces* $z=f(x, y)$ *along the* x*-axis, where* f *is given by the list*

$$x, \quad x^2, \quad x^3+xy, \quad x^3\pm xy^2, \quad x^3+xy^3, \quad x^4+xy, \quad x^4+x^2y+xy^2,$$

$$x^5\pm x^3y+xy, \quad x^3\pm xy^4, \quad x^4+x^2y+xy^3, \quad x^5+xy.$$

Here the projections are considered as the diagrams $V\to E\to B$ consisting of imbeddings and fibrations, and the equivalences are 3×2-diagrams, whose verticals are diffeomorphisms.

The singularities of a projection from a generic center are only Whitney folds and cusps (one sees a cusp along every asymptotic ray). Other singularities require special points of view. The finiteness of the list of normal forms of projections (and hence that of the list of visible contours) is not evident a priori, because there exists a continuum of nonequivalent singularities in generic 3-parameter families of projections of surfaces to the plane.

The hierarchy of tangencies may become more transparent in terms of the symplectic and contact geometries. Melrose (1976) remarked that the tangent ray geometry of a surface in Euclidean space depends on two hypersurfaces in the symplectic phase space: the first describes the metric and the second — the surface.

The same pair of hypersurfaces describes the hierarchy of asymptotic tangents. Thus we are able to transfer a large part of the geometry on the usual space surfaces to the general case of arbitrary hypersurface pairs in symplectic or contact spaces, using the geometrical intuition of the surface theory for the study of general variational problems with one-sided phase constraints.

Let Y and Z be two hypersurfaces in symplectic space X, intersecting transversally along a submanifold W. Projecting Y and Z onto their characteristics' manifolds U, V, we obtain a hexagonal diagram

$$\begin{tikzcd}
 & X & \\
Y \arrow[ur, hook] \arrow[dd] & & Z \arrow[ul, hook'] \arrow[dd] \\
 & W \arrow[uu, hook] \arrow[ul, hook'] \arrow[ur, hook] \arrow[dl] \arrow[dr] & \\
U & & V \\
 & \Sigma \arrow[uu, hook] \arrow[ul, hook'] \arrow[ur, hook] &
\end{tikzcd}$$

where Σ is the (common) manifold of critical points of the projections from W to U and to V.

Example. Let $X = \{q, p\}$ be the phase space of a Euclidean free particle (q is the particle position, p — its momentum); Y — the manifold of unit vectors ($p^2 = 1$); Z — the manifold of boundary vectors (q belongs to a hypersurface Γ). Then U is the ray space, V is Γ's tangent bundle space, W — the bundle space of the boundary (not necessarily tangent) unit vectors and Σ — the spherical tangent bundle space.

Singularities of both projections $W \to U$ and $W \to V$ at a nonasymptotic tangent unit vector are Whitney folds. Each projection defines an involution on W which is the identity on Σ.

Example. We have defined two involutions σ and τ on the manifold W of boundary unit vectors of a convex plane curve (Fig. 9). The product of involutions is the Birkhoff (1927) billiards transformation.

Melrose used the involution pairs to reduce the symplectic space hypersurface pairs to a local normal form by a C^∞-symplectomorphism (in the analytic case the series obtained are generically divergent, as is the case in the Ecalle (1975) and Voronin (1981) theories of dynamical systems at resonances).

At more complicate singularities (for instance, at asymptotic unit vectors) the symplectic space hypersurface pairs have moduli. However, one can reduce the pair formed by the first hypersurface and the intersection to simple normal form (at least formally), for the first two degeneracies of the fold. Thus we can study the singularities of the mapping which associates the ray to a boundary unit vector at the generic asymptotic and biasymptotic unit vectors.

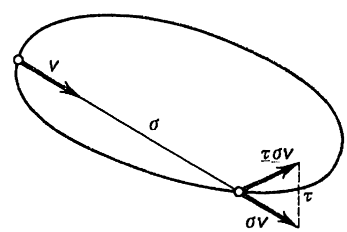


Fig. 9

The variety of critical values of this mapping is locally diffeomorphic to the product of the usual swallow-tail with a linear space. This variety lies in the symplectic space of straight lines in a standard manner:

THEOREM (1981). *All the generic symplectic structures at the point of the critical variety described above are locally reducible one into another by a critical-variety-preserving formal diffeomorphism.*

At a biasymptotic ray the variety of tangent rays is locally diffeomorphic to the product of a swallow-tail with a line. So the above theorem describes the symplectic geometry of the variety of tangent rays.

5. The obstacle problem

Consider an obstacle bounded by a smooth surface in Euclidean space. The *obstacle problem* requires a study of the singularities of the shortest path lenght from a point in the space to a fixed initial set, among paths avoiding the obstacle. This simple variational problem on a manifold with boundary is unsolved even for generic obstacles in 3-space.

The shortest path consists of segments of straight lines and of geodesics on the obstacle surfaces (Fig. 10). Hence let us consider the system of geodesics orthogonal to a fixed front. The system of all rays tangent to these geodesics is a Lagrangian variety in the symplectic space of all rays (as is every system of extremals of a variational problem). In the usual variational problems on manifolds without boundary the relevant Lagrangian variety is smooth (even in the presence of caustics). In the obstacle problem it may acquire singularities. The above theorem implies the following

COROLLARY (1981). *The Lagrangian variety in the generic obstacle problem has a semicubical cuspidal edge at the generic asymptotic rays and an "open swallow-tail" singularity at the bi-asymptotic rays.*

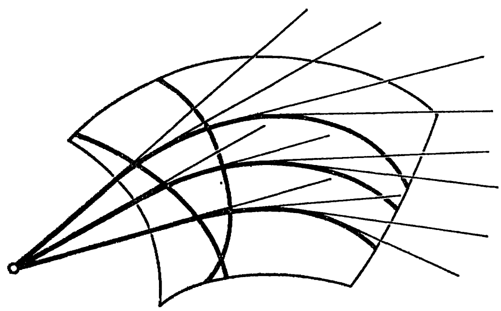


Fig. 10

The *open swallow-tail* is the surface in 4-space $\{x^5+Ax^3+Bx^2+Cx+D\}$ consisting of polynomials with at least triple roots. The differentiation of polynomials maps the open swallow-tail onto the usual one. The opening of the swallow-tail eliminates the selfinetersections but preserves the cuspidal edge (Fig. 11).

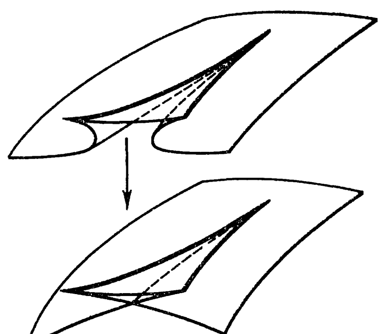


Fig. 11

THEOREM (1981). *The cuspidal edges of the wave fronts moving generically in 3-space form an open swallow-tail in the space-time (over the usual swallow-tail of the caustic).*

THEOREM (O. P. Shcherbak, 1982). *Consider a generic one-parameter family of space curves and suppose that for a given value of the parameter (of the time) the family curve has a biflatness point (of type 1, 2, 5). Then the projective dual curves form a surface in the space-time, which is locally diffeomorphic to the open swallow-tail.*

The open swallow-tail is a first representative of a large series of singularities. Consider the set of polynomials with a root of fixed comultiplicity k, $\left((x-a)^{n-k}(x^k+\ldots)\right)$ in the space of polynomials $x^n+\lambda_1 x^{n-2}+\ldots+\lambda_{n-1}$. The differentiation of polynomials preserves the comultiplicities of the roots.

THEOREM (A. B. Givental, 1981). *The sequence of varieties of polynomials with roots of fixed comultiplicity stabilizes as the degree n increases, starting with $n = 2k+1$ (i.e., at the moment of the dissociation of the self-intersections).*

Example. The open swallow-tail is the first stable variety over the usual swallow-tail.

The following Givental theory of triads (1982) formalizes the appearance of the open swallow-tail in the obstacle problem.

DEFINITION. A *symplectic triad* (H, L, l) consists of a smooth hypersurt face H in a symplectic manifold, and of a Lagrangian manifold L tangento H (with first order tangency) along a Lagrangian manifold hypersurface l.

The *Lagrangian variety generated by the triad* is the image of l in the manifold of characteristics of H.

Example 1. In the obstacle problem with boundary $\Gamma \subset \boldsymbol{R}^n$ let us consider the distance, along the geodesics of Γ, to the initial front as a function $s\colon \Gamma \to \boldsymbol{R}$. The manifold L of all the extensions of the 1-forms ds from Γ to $\boldsymbol{R}^n$ forms a triad together with the hypersurface $H\colon p^2 = 1$.

This triad generates precisely the variety of rays tangent to the geodesics of our system of extremals on Γ.

Example 2. Consider the symplectic space of polynomials $\mathscr{F} = x^d+\lambda_1 x^{d-1}+\ldots+\lambda_d$ of an even degree $d = 2m$. The polynomials, divisible by x^m, form a Lagrangian submanifold L. Let h be the Hamiltonian function of the shifts along the x axis. (This polynomial in λ is

$$h = \sum(-1)^i \mathscr{F}^{(i)}\mathscr{F}^{(j)}, \quad i+j = d, \quad \mathscr{F}^{(i)} = d^i\mathscr{F}/dx^i.)$$

The hypersurface $h = 0$ is tangent to the Lagrangian manifold L along its hypersurface l of polynomials divisible by x^{m+1} and forms with them a triad.

The variety generated by this triad is the Lagrangian open swallow-tail of dimension $m-1$ (the set of polynomials $x^{d-1}+a_1x^{d+3}+\ldots+a_{d-2}$ having a root of larger multiplicity than half the degree).

THEOREM (A. B. Givental, 1982). *The triad of Example 2 is stable. Germs of generic triads at all points are symplectically equivalent to those of Example 2.*

COROLLARY. *The variety of rays, tangent to the geodesics of the system of extremals in the generic obstacle problem is locally symplectically equivalent to the Lagrangian open swallow-tail.*

In contact geometry two sorts of Legendrian varieties are associated to the obstacle problem: the varieties of the contact elements of the fronts and the varieties of 1-jets of multi-valued time functions. The varieties of the first type ar egenerically the Legendrian open swallow-tails (they are diffeomorphically lifted Lagrangian swallow-tails). The varieties of the second type are the cylinders over the former.

Example. Consider the obstacle bounded by a plane curve with an ordinary inflection point. The fronts are the curve evolvents. They have two singularities: a usual cusp (of order 3/2) at the boundary curve of the obstacle and a 5/2-singularity at its inflectional tangent (Fig. 12). The

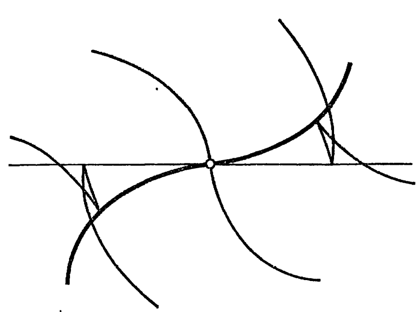


Fig. 12

Legendrian variety is nonsingular over the generic points of the obstacle curve, but over the inflectional tangent points the Legendrian variety has a cuspidal edge of order 3/2.

Let us consider the 3-space of the plane contact elements (fibered over the plane). All contact elements of all the evolvents of a generic curve form a surface in this 3-space. Let us consider the 3-space of polynomials x^3+ax^2+bx+c (fibered over the plane of their derivatives). All those polynomials having multiple roots form a surface in this 3-space.

THEOREM (1978). *The germ of the first surface at the tangent at an inflection point of a generic curve is diffeomorphic to the germ of the second surface at zero, by a fibre-preserving diffeomorphism.*

This surface (Fig. 13), together with the $c = 0$ surface representing the plane contact elements at the obstacle boundary points, forms the variety of singular orbits for the reflection group B_3. This remark has led to the boundary singularity theory (1978), of which I shall only mention the following.

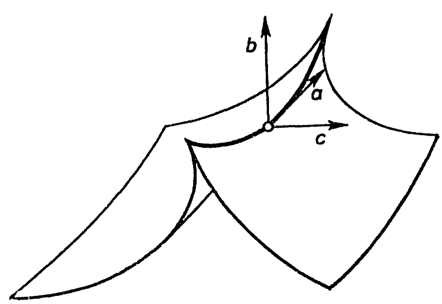


Fig. 13

Example (I. G. Shcherbak, 1982). Consider a generic curve on a generic surface in Euclidean 3-space. At some points the curve touches the surface curvature line. The boundary Lagrangian singularity theory implies that this situation is governed by the exceptional Weyl group F_4: the union of the focal sets of the surface and of the curve with all the surface normals at the points of the curve forms a variety which is locally diffeomorphic to the F_4 caustic.

The boundary Lagrangian singularity theory implies an amusing "Lagrange duality", which interchanges the singularity of a function on the ambient space with that of its restriction to the boundary: this duality is a modernized version of the "Lagrange multipliers rule" (I. G. Shcherbak, 1982).

Returning to an inflection point of a plane curve, consider the graph of the (multi-valued) time function for the obstacle problem. The level sets of this time function are the obstacle evolvents. Hence the graph has the shape drawn in Fig. 14 (1978); this surface has two cuspidal edges (of orders 3/2 and 5/2).

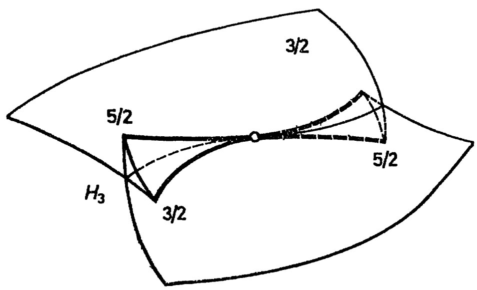


Fig. 14

When I showed this surface to A. B. Givental (1982), he recognized the singular orbit variety of the group H_3 of symmetries of the icosahedron drawn by O. V. Lyashko (1981). Givental's conjecture was rapidly confirmed:

THEOREM (O. P. Shcherbak, 1982). *The graph of the (multi-valued) time function in the generic plane obstacle problem is diffeomorphic to the variety of singular orbits at the inflection points of the obstacle boundary.*

THEOREM[1] (O. V. Lyashko, 1981). *The variety of singular orbits of H_3 is diffeomorphic of the space of polynomials $x^5+ax^4+bx^2+c$ having a multiple root.*

Lyashko's theorem describes the variety of singular orbits of H_3 as the union of the tangents to the curve (t, t^3, t^5) in 3-space, while Shcherbak deals with any curve $(t+\ldots, t^3+\ldots, t^5+\ldots)$.

A generic front in the 3-space obstacle problem must have a singularity of the same type at the point of tangency of an asymptotic ray with the obstacle surface.

In this paper I have not even mentioned many important aspects of the Lagrangian and Legendrian singularity theory, especially the global ones, such as the theory of the coexistence of singularities (the Lagrangian and Legendrian cobordism theories reduced to homotopy problems by Ya. M. Eliashberg, the Lagrangian and Legendrian characteristic

[1] A similar description of $\Sigma(H_4)$ was found by O. P. Shcherbak in 1983: it is based on an inclusion of the H_4 graded local algebra defined by the invariants convolution, into the E_8 graded local algebra: $x^3+y^5+axy^3+by^3+cx+d$.

classes of V. A. Vasiliev, which are generalizations to higher singularities of V. P. Maslov's class, and so on).

I have not even mentioned the extensive classification of the simple projections (Goryunov, 1981), the theory in which, for instance, the exceptional root system F_4 is an ancestor of a whole family of descendants F_μ. One can find details of those theories and the extensive relevant bibliography in the surveys [1and [2].

In spite of the progress of the ray system geometry during the past three centuries, from Huygens up to now, the drawing of pictures very similar to those one finds in Huygens' works is still one of the main sources of new discoveries in this difficult domain where even the 3-dimensional problem is still unsolved and where numerous useful but unexpected interrelations with other branches of mathematics (such as relation of the obstacle problem to the group H_3 of symmetries of the icosahedron) still remain mysterious.

References

[1] Arnold V. I., Singularities of Rays Systems, *Uspekhi Mat. Nauk* **38** (2) (1983).
[2] Arnold V. I., Singularities in the Calculus of Variations, *Itogi Nauki, Sovremennye problemy mat.* **22**, VINITI, Moscow, 1983.
[3] Arnold V. I., *Catastrophes Theory*, Springer, 1983.

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

P. ERDÖS

Extremal Problems in Number Theory, Combinatorics And Geometry

During my long life I wrote many papers on these subjects [1]. There are many fascinating and difficult unsolved problems in all three of these topics. I have to organize the problems in some order. This is not an easy task and anyway not one of my strong points.

In number theory I will mainly discuss questions related to van der Waerden's theorem on long arithmetic progressions and problems in additive number theory.

In geometry the questions I want to discuss are either metrical problems, e.g. number of distinct distances which must occur between points in a metric space. The metric space usually will be our familiar E_2. I will also discuss incidence problems of points in E_2. These problems have a purely combinatorial interpretation too, but the results in E_2 are completely different than in the finite geometries.

In combinatorics I will discuss Sperner, Ramsey and Turán type problems and will try to emphasize their applications to number theory and geometry.

Since I must, after all, remain myself, I can not entirely refrain from stating some old and new problems which, in my opinion, perhaps have been undeservedly neglected.

I hope the reader will forgive a very old man for some personal and historical reminiscences but to save space I will try to write only facts which I did not mention elsewhere.

1. Number theory

First I discuss problems in number theory. Here some of the most striking and significant questions are those connected with the results of van der Waerden and Szemerédi.

Van der Waerden proved more than 50 years ago that if we partition the integers into two classes, at least one of them contains an arbitrarily large arithmetic progression. Many beautiful and important extensions and modifications are known, e.g. the Hales–Jewett theorem and Hindman's theorem but we have no space to discuss these here. A very nice book on this subject has been published recently by Graham, Rothschild and Spencer [2]. (This book contains a very extensive list of references and the references which I suppress here can be found there.) The finite form of van der Waerden's theorem is:

Let $f(n)$ be the smallest integer for which if we divide the integers not exceeding $f(n)$ into two classes, then at least one of them contains an arithmetic progression of n terms. Van der Waerden's original proof gives an explicit upper bound for $f(n)$ but his bound increases very fast: in fact as fast as the well known Ackerman function (which increases so fast that it is not primitive recursive). The best lower bound due to Berlekamp, Lovász and myself increases only exponentially, like a power of 2. The first task would be to prove (or disprove) that $f(n)^{1/n}$ tends to infinity but that $f(n)$ tends to infinity more slowly than Ackerman's function.

Another equally important task would be to find a sharpening of Szemerédi's theorem: Denote by $r_k(n)$ the smallest integer ϱ for which every sequence $a_1 < \ldots < a_\varrho \leqslant n$ contains an arithmetic progression of k terms. Turán and I conjectured 50 years ago that for every k, $r_k(n) = o(n)$. This conjecture was proved for $k = 3$ by K. F. Roth, then later by Szemerédi for $k = 4$ and finally by Szemerédi for every k. A few years ago Fürstenberg proved Szemerédi's theorem by methods of ergodic theory. This proof does not give an explicit upper bound for $r_k(n)$. Fürstenberg and Katznelson proved the n-dimensional generalization of Szemerédi's theorem for which there is so far no other proof. It is not yet possible to tell the potentialities and possible limitations of this new method [3].

K. F. Roth and F. Behrend proved that

$$\frac{n}{e^{c_1\sqrt{\log n}}} < r_3(n) < \frac{c_2 n}{\log\log n}. \tag{1}$$

No useful upper estimation for $r_k(n)$ is known for $k > 3$. Szemerédi and I observed that it is not even known whether $r_k(n)/r_{k+1}(n) \to 0$. It would be very desirable to improve the upper and lower bounds in (1) and to obtain some useful upper bounds for $r_k(n)$. In particular, is it true that

$$r_k(n) < \frac{n}{(\log n)^l} \tag{2}$$

for every k and l if $n > n_0(k, l)$? I offered a reward of \$ 3,000 for a proof or disproof of (2). (2), if true, would of course imply my old conjecture: If $\sum \frac{1}{a_i} = \infty$, then for every k there are k a's forming an arithmetic progression. This in turn would imply that there are arbitrarily long arithmetic progressions in the primes. Recently 18 primes in an arithmetic progression were found by Pritchard [33]. It seems certain that a much stronger result holds for primes: For every k there are k consecutive primes forming arithmetic progression. But this problem certainly cannot be attacked by any of our present-day methods, and is in fact beyond any methods likely to be at our disposal in the near or distant future. Schinzel's well-known hypothesis H would imply it. Van der Corput, Estermann and Tchudakoff independently deduced by Vinogradov's method that the number of even numbers $2n < x$ which are not the sums of two primes in many ways is less than $x/(\log x)^d$ for every d if $x > x_0(d)$. This was later improved by Montgomery and Vaughan to x^{1-c}. (In fact by Goldbach's conjecture all even numbers but 2 are sums of two primes.) These results immediately imply that there are infinitely many triples of primes in an arithmetic progression. It is not yet known whether there are infinitely many quadruples of primes in an arithmetic progression.

An old conjecture of mine (in fact one of my first conjectures, which perhaps did not receive as much attention as it deserved), can be stated as follows: Let $\{f(n)\}$ be an arbitrary sequence with $f(n) = \pm 1$. Then for every $c > 0$ there are an m and a d such that

$$\left|\sum_{k=1}^{m} f(kd)\right| > c. \tag{3}$$

Note that I permit fewer arithmetic progressions here than in van der Waerden's theorem but I also ask for much less. A weaker variant of (3) states that if $f(n) = \pm 1$ and $f(n)$ is multiplicative, then

$$\overline{\lim_{n\to\infty}} \left|\sum_{k=1}^{n} f(k)\right| = \infty. \tag{4}$$

Tchudakoff [4] stated independently in another context a more general conjecture. Here I just would like to call attention to a large class of problems called problems on irregularities of distributions or discrepancy problems. The first results in this subject were found by van der Corput and Aardenne-Ehrenfest and later by K. F. Roth and W. Schmidt, also by Spencer and myself. Recently very striking results were obtained by J. Beck [32]. Very recently these problems were discussed by V. T. Sós in a more general setting. Her paper [30] will appear soon. Here I restrict myself to problems related to van der Waerden's theorem. Denote by $f(n; l)$ the smallest integer for which if we divide the integers not exceeding $f(n; l)$ into two classes, then there is an arithmetic progression of n terms which contains at least $n/2+l$ terms in one of the classes. $f(n; n/2)$ is our old $f(n)$. I easily proved by the probability method that for $l > \varepsilon n$

$$f(n; l) > (1+c_\varepsilon)^n. \tag{5}$$

For small values of ε, (5) is perhaps not very far from being best possibler It would be very interesting and useful to obtain good upper and lowe. bounds for $f(n; l)$. It would be especially interesting if one could determine the dependence of $f(n; l)$ on l. The trouble is that there are no non-trivial upper bounds known for $f(n; l)$, not even if l is bounded. As far as I know the only result of this kind is due to J. Spencer who determined $f(2n; 1)$, i.e. he determined the smallest integer $f(2n; 1)$ for which one cannot divide the integers $1, 2, \ldots, f(2n; 1)$ into two classes so that every arithmetic progression of $2n$ terms contains precisely n terms from each class [5].

2. Combinatorics and additive number theory

Now I discuss problems in combinatorial additive number theory. For a fuller history and discussion of such problems I refer the reader to the excellent book [6] by Halberstam and Roth. Perhaps my oldest conjecture (more than 50 years old) is the following: Denote by A a sequence $1 \leqslant a_1 < \ldots < a_k \leqslant n$ of integers. Assume that all the sums $\sum_{l=1}^{k} \varepsilon_i a_i$ ($\varepsilon_1 = 0$ or 1) are distinct. Is it true that $k = \log n/\log 2 + C$ for some absolute constant C? The powers of 2 show that $C \geqslant 1$. L. Moser raised this problem independently. Moser and I proved that

$$k < \frac{\log n}{\log 2} + \frac{\log\log n}{2\log 2} + O(1).$$

Conway and Guy [7] showed that $C \geqslant 2$. They found 24 integers not exceeding 2^{22} for which all the subset sums are different. It has been conjectured that their construction is perhaps optimal and that $C = 3$.

Now I discuss some problems of Sidon. Sidon called a finite or infinite sequence A a $B_k^{(r)}$ sequence if the number of representations of every integer n as the sum of k or fewer a's is at most r. Let us first assume $k = 2$, $r = 1$, i.e. we assume that the sums $a_i + a_j$ are all distinct. First we consider infinite B_2 sequences (for $r = 1$ a $B_k^{(1)}$ sequence will be denoted by B_k). Sidon asked: Determine the slowest possible growth of a B_2 sequence. The greedy algorithm immediately gives that there is a B_2 sequence for which $a_k < ck^3$. On the other hand I proved that for every B_2 sequence

$$\limsup_{k\to\infty} \frac{a_k}{k^2(\log k)} \geqslant 1. \tag{6}$$

I have been able to improve (6). Further I proved that there is a B_2 sequence for which

$$\liminf \frac{a_k}{k^2} \geqslant \frac{1}{2}. \tag{7}$$

Krückeberg replaced $\frac{1}{2}$ in (6) by $1/\sqrt{2}$ and I conjectured that $1/\sqrt{2}$ can further be improved to 1 which, if true, would be best possible. I could prove (7) if I could prove that if $a_1 < a_2 < \ldots < a_l$ is a B_2 sequence, then it can be embedded into a B_2 sequence $a_1 < \ldots < a_l < a_{l+1} < \ldots < a_k$ with $a_k = (1+o(1))k^2$. Perhaps the following stronger result holds: Every B_2 sequence can be embedded into a perfect difference set. Rényi and I proved by probabilistic methods that to every $\varepsilon > 0$ there is an $r = r(\varepsilon)$ for which there is a $B_2^{(r)}$ sequence satisfying $a_k < k^{2+\varepsilon}$ for every k. I would expect that in fact there is a B_2 sequence satisfying $a_k < k^{2+\varepsilon}$, but the proof of this is nowhere in sight. In fact $a_k < ck^3$ remained unimproved for nearly 50 years. Recently Ajtai, Komlós and Szemerédi [8] proved by a novel and very ingenious combinatorial method that there is a B_2 sequence for which

$$a_k < \frac{ck^3}{\log k}.$$

This new method was recently applied by Komlós, Pintz and Szemerédi to Heilbronn's problem [9].

Denote by $f_k(x)$ the largest integer for which there is a B_k sequence having $f_k(x)$ terms not exceeding x. Turán and I proved that

$$\big(1+o(1)\big)x^{1/2} < f_2(x) < x^{1/2}+cx^{1/4}. \tag{8}$$

Lindstrom proved $f_2(x) < x^{1/2}+x^{1/4}+1$ which at present is the best upper bound for $f_2(x)$.

The lower bound of (8) was also proved by Chowla. Turán and I conjectured that

$$f_2(x) = x^{1/2}+O(1). \tag{9}$$

(9), if true, is probably very difficult. Bose and Chowla proved that for every k

$$f_k(x) \geqslant \big(1+o(1)\big)x^{1/k}.$$

Bose observed that our method with Turán fails to give $f_k(x) \leqslant \big(1+o(1)\big)x^{1/k}$ and in fact this problem is still open. In fact I could never prove that if A is infinite B_k sequence ($k>2$), then

$$\limsup a_l/l^k = \infty.$$

Sidon asked me more than 50 years ago: Denote by $f(n)$ the number of solutions of $n = a_i+a_j$. Is there a basis of order 2 (i.e. every integer is the sum of two a's) for which $f(n) = o(n^\varepsilon)$, for every $\varepsilon > 0$? By probabilistic methods I proved that there is a basis of order 2 for which

$$c_1 \log n < f(n) < c_2 \log n$$

which is very much stronger than Sidon's conjecture. Turán and I further conjectured that for every basis of order 2 we have

$$\overline{\lim_{n\to\infty}} f(n) = \infty. \tag{10}$$

Perhaps (10) already follows if we assume only that $a_k < ck^2$ holds for some c and every k.

Is there a basis of order 2 for which

$$f(n)/\log n \to 1? \tag{11}$$

Probably (11) will not be quite easy, since (unless I overlook an obvious idea) the probability method does not seem to help with (11).

D. Newman and I conjectured that there is a $B_2^{(2)}$ sequence which is not the union of a finite number of B_2 sequences. Three years ago I found a very simple proof of this conjecture [10]. Nesetril and Rödl proved the related conjecture for $B_2^{(r)}$ sequences. In fact I proved that there is a $B_2^{(2)}$ sequence having n^3 terms no subsequence of which having more than $2n^2$ terms is a B_2 sequence. To see this consider the $B_2^{(2)}$ sequence

$$4^u+4^v, \quad 1 \leqslant u \leqslant n, \; n < v \leqslant n+n^2.$$

This is a $B_2^{(2)}$ sequence of n^3 terms and no subsequence having more than $2n^2$ terms can be a B_2 sequence. To see this consider a complete bipartite graph of n black and n^2 white vertices. The black vertices correspond to the numbers 4^u and the white vertices to 4^v. 4^u+4^v corresponds to the edge joining 4^u and 4^v. A simple graph theoretic argument shows that every subgraph of $2n^2$ edges contains a C_4, i.e. a rectangle. This shows that the subsequence is not a B_2 sequence. Observe that $n^2 = (n^3)^{2/3}$. I cannot decide if the exponent $\frac{2}{3}$ is best possible. Perhaps it could be improved to $\frac{1}{2}$ but I doubt it [11]. V. T. Sós and I considered $B_2'^{(r)}$ sequences, i.e. sequences $a_1, \ldots, a_l, \ldots$, where the number of solutions of $m = a_i - a_j$ is at most r. Of course, for $r=1$ we obtain our familiar B_2 sequences. We could not decide whether there is a $B_2'^{(r)}$ sequence which is not the union of a finite number of B_2 sequences. It is easy to construct a B_2 sequence for which every integer has a unique representation $a_i - a_j$. On the other hand it is easy to see that if $a_k < \lambda k^2$, then the number of solutions of $a_j - a_i = t$ cannot be bounded. We plan to write a paper at a later date on these and other problems on $B_2'^{(r)}$ sequences.

To complete this chapter I state two unsolved problems: Let $\varphi(n) \to 0$ and $1 \leqslant a_1 < a_2 < \ldots < a_x \leqslant n$ be the largest set of integers for which the number of distinct integers of the form $a_i + a_j$ is $\leqslant \big(1+\varphi(n)\big)\binom{n}{2}$. I can prove that $x \geqslant \big(1+o(1)\big)\dfrac{2}{3^{1/2}}\, n^{1/2}$, and hope that for some constant $\varepsilon > 0$, $x < (1-\varepsilon)\,(2n)^{1/2}$. This, if true, would imply that a harmonious graph of n vertices must have fewer than $(1-\eta)\binom{n}{2}$ edges [12] for some constant $\eta > 0$.

Silverman and I asked: Let $h(n)$ be the largest integer for which there is a sequence $0 < a_1 < \ldots < a_{h(n)} \leqslant n$ so that none of the sums $a_i + a_j$ is a square. How large is $h(n)$? This harmless looking question leads to surprising complications [13].

3. Geometrical problems, global results

Next I discuss geometrical problems. Let $x_1, \ldots, x_n$ be distinct points in E_k. Denote by $d(x_i, x_j)$ the distance between x_i and x_j. Let t denote the number of distinct distances and let $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_t$ denote the multiplicities of the distances $\left(\sum_{i=1}^{t} a_i = \binom{n}{2}\right)$. We will mainly study the maximum possible value of a_1 and the minimal value of t. We will study the problems both globally and locally, where "locally" means that we study the distribution of "distances from one point x_i". We will study these problems both if the points are in general position and also if they are restricted in various ways, e.g. they form a convex set or no three are on a line. It will turn out that all these problems lead to various interesting and difficult questions and we are far from their final solutions. Many of them lead to interesting combinatorial extremal problems. G. Purdy and I hope to finish a book on these geometrical problems before this decade is over. We have written several joint papers on this subject.

V. T. Sós and I tried to obtain conditions $\left(\text{other than the trivial } \sum_{i=1}^{t} a_i = \binom{n}{2}\right)$ which would permit us to embed the points in E_k. This question is nontrivial even for $k = 1$ and we have obtained only preliminary results (in many cases with the help of various colleagues). We hope to return to these problems elsewhere. I was told of the following attractive conjecture of Specker. It is easy to see that for every choice of the multiplicities $a_1, \ldots, a_t$ the points can be embedded into E_k for some $k \leqslant n-1$. Specker conjectured that $k = n-1$ is needed only for the regular simplex, i.e., if $a_1 = \binom{n}{2}$. I have never looked seriously at this nice conjecture but I am told that it does not seem to be trivial. V. T. Sós and I raised the following question: Is there an $f(k)$ so that if the n points are in E_k and the minimum multiplicity $a_t > n$ then $n < f(k)$.

Now let us return to our subject. First of all I wish to remark that both on the metrical and on the incidence problems important progress has been made in the last 2 years by J. Beck, J. Spencer, F. Chung, E. Szemerédi and W. Trotter. I have to apologize that I mainly restrict myself to my own problems (not because I consider them more important but because I know more about them). I just want to remark that very recently P. Ungár [31] completely solved a problem of Scott by proving that n points in E_2 determine at least $n-1$ distinct directions (sharpening an earlier result of Burnet and Purdy).

Denote by $f_k(n)$ the largest possible value of a_1 and $g_k(n)$ the smallest possible value of t. Probably $k = 2$ is the most interesting and difficult case. For $k = 1$ everything is trivial: $f_1(n) = g_1(n) = n-1$. For $k > 2$ interésting problems remain, but to save space I will only refer to the literature.

I observed in 1945 that

$$n^{1+c/\log\log n} < f_2(n) < c_2 n^{3/2} \tag{12}$$

and conjectured that the lower bound in (12) is best possible or at least not far from being best possible. The lower bound is given by the triangular or square lattice and perhaps some sort of lattice gives the true lower bound. V. T. Sós and I conjectured that the n points which give $f_2(n)$ must contain an equilateral triangle or a square or at least a set of 4 points which determine at most 2 (or perhaps 3) distinct distances. Further we asked: Is it true that $f_2(n) - a_2 \to \infty$ as $n \to \infty$? Is it true that the configurations which maximize a_1 are the same which minimize t? The answer is almost certainly no. Join two points if their distance is 1. Assume that the distance 1 occurs $f_2(n)$ times. We could get no useful properties of this graph. Of course, it must be connected. It is easy to see that this graph cannot contain a $K(2, 3)$ and this was the way I originally proved $f_2(n) < cn^{3/2}$. With a little trouble we could enumerate all the forbidden subgraphs having fewer than k vertices, as long as k is not too large. Once I hoped that the exclusion of finitely many of these forbidden graphs will give $f_2(n) < n^{1+\varepsilon}$. But now I rather believe that for $n > n_0(k)$ there is a graph of n vertices and $cn^{3/2}$ edges not containing any of the forbidden graphs having $\leqslant k$ vertices.

Szemerédi proved 10 years ago that $f_2(n) = o(n^{3/2})$. Two years ago Beck and Spencer proved $f_2(n) < n^{3/2-\varepsilon}$ for some $\varepsilon > 0$. This was improved by Fan Chung, Szemerédi, Trotter and Spencer to $f_2(n) < n^{4/3}$. Unfortunately their method does not seem to give $f_2(n) < n^{1+\varepsilon}$. I also observed in 1946 that $g_2(n) > \sqrt{n-1} - 1$. This was improved by L. Moser to $g_2(n) > cn^{2/3}$ and last year Fan Chung proved $g_2(n) > cn^{5/7}$. This has also been improved to $g_2(n) > cn^{3/4}$.

4. Distance distribution, local results

I conjectured that if $x_1, x_2, \ldots, x_n \in E_2$ and one denotes by $t_i(n)$ the number of different distances from x_i then

$$\max_i t_i(n) > cn/(\log n)^{1/2}.$$

Beck proved $\max_i t_i(n) > cn^{5/7}$ and this was also improved to $\max_i t_i(n) > n^{3/4}$. In fact I conjectured that

$$\sum_{i=1}^{n} t_i(n) > cn^2/(\log n)^{1/2}. \tag{13}$$

Perhaps (13) is a bit too optimistic but as far as I know no counterexample is known.

I conjectured that for any n points $x_1, \ldots, x_n$ in the plane there is an x_i so that the number of points equidistant from x_i is $o(n^\varepsilon)$ and perhaps it is less than $n^{c/\log\log n}$. The lattice points again show that (if true) this conjecture is best possible. It is trivial that this result holds with $cn^{1/2}$ instead of n^ε and recently Beck proved it with $o(n^{1/2})$. Denote by $a_i(n)$ the largest number of points equidistant from x_i. The most optimistic conjecture is that

$$\sum_{i=1}^{n} a_i(n) < n^{1+c/\log\log n}. \tag{14}$$

Again, perhaps (14) is a bit too optimistic.

5. Distance distributions with conditions

Below we shall assume some additional properties $\mathscr{P}$ of the points $x_1, \ldots, x_n$, and denote the corresponding functions by $f_k(\mathscr{P}, n)$, $g_k(\mathscr{P}, n)$, $t_i(\mathscr{P}, n)$. Let $\mathscr{C}$ denote that $x_1, \ldots, x_n$ form a convex set. I conjectured and Altman proved that

$$g_2(C, n) = \left[\frac{n}{2}\right]. \tag{15}$$

Szemerédi conjectured that (15) remains true if we assume only that no three of the points are on a line, but his proof gives only $g_2(\mathscr{L}, n) \geqslant \left[\frac{n}{3}\right]$ (where $\mathscr{L}$ denotes the above property). L. Moser and I conjectured that

$$f_2(C, 3n+1) = 5n$$

but we could not even prove $f_2(C, n) < cn$. I also conjectured that for a convex set

$$\max_i t_i(C, n) \geqslant \left[\frac{n}{2}\right] \tag{16}$$

but (16) is still open. Perhaps (16) remains true if we assume only that no three of the x_i's are on a line: $\max_i t_i(\mathscr{L}, n) \geqslant [n/2]$. I further conjectured that in the convex case there is always an x_i so that no three of the other vertices are equidistant from it. This was disproved by Danzer but perhaps remains true if 3 is replaced by 4. Here again convexity could be replaced by the condition $\mathscr{L}$.

Let $\mathscr{L}^*$ denote that no three points are on a line, no four on a circle. Is it then true that

$$g_2(\mathscr{L}^*, n)/n \to \infty ?$$

I could not even exclude $g_2(\mathscr{L}^*, n) > cn^2$, but perhaps here I overlook an obvious argument. I could not exclude the possibility that $g_2(\mathscr{L}^*, n) = n-1$ and $a_i = i$, $1 \leqslant i \leqslant n-1$. I thought that this is impossible for $n \geqslant 5$, but colleagues found not quite trivial examples for $n = 5$ and $n = 6$.

Let $\mathscr{F}$ denote that every set of 4 points determines at least 5 distinct distances. Is it then true that $g_2(\mathscr{F}, n) > cn^2$? Is the chromatic number of the hypergraph formed by the quadruples determining $\geqslant 5$ distances bounded? If the points are on a line, this chromatic number is 2.

I could not prove $g_2(\mathscr{F}, n) > cn^2$ even if we assume that every set of 5 points determine at least 9 distinct distances.

Below we delete $\mathscr{P}$ from our notation. Very likely, if our set contains no isosceles triple (i.e., if every set of three points determines three distinct distances) then $g_2(n)/n \to \infty$.

Assume finally that the points are on a line and that every set of 4 points determines at least 4 distinct distances. Then $g_1(n)/n \to \infty$ but $g_1(n)$ can be less than $n^{1+\varepsilon}$. The number of these problems could clearly be continued but it is high time to stop.

6. Incidence problems

Now I discuss incidence problems. Sylvester conjectured and Gallai proved that if n points are given in E_2, not all on a line, then there is always a line which goes through exactly two of the points. The finite geometries show that special properties of the plane (or E_k) must be used here. Motzkin conjectured that for $n > n_0$ there are at least $[n/2]$ such lines. He further observed that, if true, this conjecture is best possible for infinitely many n. Hansen recently proved this conjecture, sharpening a previous result of W. Moser and L. M. Kelly. Hansen's proof has not yet been published.

Croft, Purdy and I conjectured that if $x_1, \ldots, x_n$ is any set of n points in E_2, then the number of lines which contain more than l points is less than cn^2/l^3. The lattice points in the plane show that, if true, this conjecture is best possible. Szemerédi and Trotter recently proved this conjecture. Thus, in particular, there are fewer than $cn^{1/2}$ lines which contain more than $n^{1/2}$ points. The finite geometries again show that special properties of the plane have to be used.

A few weeks later Beck independently proved our conjecture by a different method but in a slightly weaker form. The strong form of our conjecture was needed to prove another conjecture of mine. Denote by $L_1, \ldots, L_m$ the lines determined by our points. By a well-known result of de Bruijn and myself $m \geqslant n$. Denote by $|L_i|$ the number of points on L_i, $|L_1| \geqslant |L_2| \geqslant \ldots \geqslant |L_m|$. I conjectured that the number of distinct sets of cardinalities $\{|L_1|, \ldots, |L_m|\}$ is between $e^{c_1 n^{1/2}}$ and $e^{c_2 n^{1/2}}$. The lower bound is easy and Szemerédi and Trotter proved the upper bound. I have a purely combinatorial conjecture. Let $|S| = n$ and let $A_i \subset S$, $1 \leqslant i \leqslant m$ be a partially balanced block design of S, i.e. every pair $\{x_i, x_j\}$ be contained in one and only one of the A's. I conjectured that the number of distinct sets of cardinalities in $\{|A_1|, \ldots, |A_m|\}$ is between $n^{c_1 n^{1/2}}$ and $n^{c_2 n^{1/2}}$. The upper bound is easy but the lower bound is still open. Rödl recently informed me that the lower bound is also easy.

About 100 years ago Sylvester asked himself the following problem. Assume that no four of the x_i's are on a line. Determine or estimate the largest number of triples of points which are on a line. $\frac{1}{3}\binom{n}{2}$ is a trivial upper bound and Sylvester proved that $n^2/6 - cn$ is possible. The best possible value of c is not yet known. Thus here the difference between the plane and block designs is not so pronounced. A few years ago Burr, Grünbaum and Sloane [14] wrote a comprehensive paper on this subject. They gave a plausible conjecture for the exact maximum. Their paper contains extensive references. Some of their proofs are simplified in a forthcoming peper of Füredi and Palásthy.

Surprisingly, an old conjecture of mine has so far been intractable. Assume $k \geqslant 4$ and that no $k+1$ of our points are on a line. Let $l_k(n)$ be the maximum number of k-tuples which are on a line. Then $l_k(n) = o(n^2)$.

B. Grünbaum proved that $l_k(n) > cn^{1+1/k-2}$ is possible and perhaps this result is best possible.

Dirac conjectured that if $x_1, \ldots, x_n$ are n points not all on a line and we join every two of them, then there is always an x_i so that at least $n/2 - c$ distinct lines go through x_i. If true, then apart from the value of c this is easily seen to be best possible. Szemerédi and Trotter and a few weeks later Beck proved this conjecture with $c_1 n$ instead of $n/2$. Finally Beck proved the following old conjecture of mine. Let there be given n points, at most $n-k$ on a line. Then these points determine at least ckn distinct lines. Perhaps the correct value of c is $\frac{1}{6}$ in any case. Beck gets a very small value of c. Many very interesting questions have completely been omitted, e.g. Borsuk's conjecture [15]. Some more geometric problems will be mentioned in the last chapter on combinatorial problems, where combinatorial theorems directly imply geometric or number-theoretic results.

7. Combinatorial problems

In this final chapter I discuss combinatorial problems. Many mathematicians, including myself, wrote several survey papers on this subject [16] and therefore I will try to keep this chapter short. Also recently appeared an excellent book of Bollobás [17] and several very interesting papers of Simonovits will soon appear. Thus, apart from a few favourite problems, I will mention only results having applications in number theory or geometry. Perhaps the first significant result in this subject is the following theorem of Sperner [18]: Let $|S| = n$, $A_i \subset S$, $1 \leqslant i \leqslant T_n$ be a family of subsets of S no one of which contains the other, then

$$\max T_n = \binom{n}{[n/2]}. \tag{17}$$

Sperner's theorem was forgotten for a long time, perhaps even by its discoverer. When I first met Sperner in Hamburg more than 20 years ago, I asked him about his result. He first thought that I asked him about his much better known lemma in dimension theory, and it turned out that he all but forgot about (17). (17) in fact was used a great deal in the theory of additive arithmetical functions. As far as I know, the first use of (17) was due to Behrend and myself. Behrend and I proved (Behrend [19] a few months earlier) that if $1 \leqslant a_1 < \ldots < a_h \leqslant x$ is a sequence of integers in which no one divides the other, then

$$\sum_{c=1}^{h} \frac{1}{a_i} < \frac{c \log x}{(\log\log x)^{1/2}}. \tag{18}$$

Pillai and both of us observed that (18) is best possible, apart from the value of c. Later Sárközy, Szemerédi and I [20] determined the best value of c in (18). It seems likely that there is no simple characterization of the extremal sequences.

Now I discuss some extremal problems on graphs and hypergraphs. As stated, many papers and the book of Bollobás have appeared recently on this subject, thus I will be very sketchy. Let $G^{(r)}$ be an r-uniform hypergraph (i.e. the basic elements of $G^{(r)}$ are its vertices and r-tuples). For $r = 2$ we obtain the ordinary graphs [21]. $G(n, e)$ will denote a graph (r-uniform hypergraph) of n vertices and e edges (r-tuples). Let $f(n; G^{(r)})$ be the smallest integer so that every $G\big(n; f(n; G^{(r)})\big)$ (i.e. every r-graph of n vertices and $f\big(n; G^{(r)}(n)\big)$ edges (i.e. r-tuples)) contains $G^{(r)}$ as a subgraph. If $r = 2$ and $G^{(2)}$ is a complete graph of l vertices $K(l)$, Turán determined more than 40 years ago $f\big(n; K(l)\big)$ for every l. He also asked for the determination of $f(n; G)$ for more general graphs. Thus started an interesting and fruitful new chapter in graph theory. In particular he asked for the determination of $f\big(n; K^{(r)}(l)\big)$ where $K^{(r)}(l)$ is the complete r-graph of l vertices. This problem is probably very difficult. It is easy to see that

$$\lim_{n\to\infty} f\big(n; K^{(r)}(l)\big) \Big/ \binom{n}{r} = c_{r,l}$$

always exists. $c_{2,l} = 1-1/(l-1)$, but for no $l > r > 2$ is the value of $c_{r,l}$ known. Turán had some plausible conjectures. One possible reason for the difficulty of this problem is that (while Turán proved the uniqueness of his extremal graphs for $r = 2$ and every l) W. G. Brown, and in more general form Kostochka, proved that for $r > 2$ there are many different extremal graphs [22]. Now, for $r = 2$ I state some of our favourite conjectures with Simonovits. It is well known that

$$f(n; C_4) = \big(\tfrac{1}{2} + o(1)\big)\, n^{3/2}. \tag{19}$$

Try to characterize the (bipartite) graphs for which

$$f(n; G) < cn^{3/2}. \tag{20}$$

(One can easily see that if $f(n, G) = o(n^2)$, then G is bipartite.) Our conjecture (perhaps more modestly it should be called a guess) is that (20) holds if any only if G is bipartite and has no subgraph each vertex of which has degree (or valency) greater than 2. Unfortunately we could neither prove the necessity nor the sufficiency of this attractive, illuminating (but perhaps misleading) conjecture. A weaker conjecture, having a better chance of being true, states: Let G satisfy (20). Define G' by adjoining a new vertex to G and join it to two vertices of G of different colour. Then G' also satisfies (20). Further, we conjectured that if G is bipartite, then there is some rational α, $1 \leqslant \alpha < 2$, for which

$$\lim f(n; G)/n^{\alpha} = c, \qquad 0 < c < \infty. \tag{21}$$

Further, for every rational $\alpha \in [1, 2)$ there is a G for which (21) holds [23]. It is well known that (21) is false for $r > 2$, but perhaps for every $G^{(r)}$

$$\lim_{n\to\infty} f(2n; G^{(r)})/f(n; G^{(r)}) = c(G^r) \tag{22}$$

exists and differs from 0 and ∞.

Nearly 50 years ago I investigated the following extremal problem in number theory: Let $1 \leqslant a_1 < a_2 < \ldots < a_h \leqslant x$. Assume that all the products $a_i a_j$ are distinct. Put $f(x) = \max a_h$: estimate $f(x)$ as accurately as possible. I proved that there are positive constants $c_2 \geqslant c_1 > 0$ such that [23]

$$\pi(x) + c_1 x^{3/4}/(\log x)^{3/2} < f(x) < \pi(x) + c_2 x^{3/4}(\log x)^{3/2} \tag{23}$$

(where $\pi(x)$ is the number of primes $\leqslant x$). The proof of (23) was based on the inequality

$$c_3 n^{3/2} < f(n; c_4) < c_4 n^{3/2}. \tag{24}$$

(24) was proved at that time by E. Klein and me.

Another number theoretic application of graph theory is as follows:

Denote by $K^{(r)}(t, \ldots, t)$ the r-graph of rt vertices $x_1^{(j)} \ldots x_r^{(j)}$, $1 \leqslant j \leqslant t$ having t^r edges $\{x_{i_1}^{(1)}, x_{i_2}^{(2)}, \ldots, x_{i_t}^{(t)}\}$. I proved (for $t = 2$ this was proved earlier in a sharper form by Kövári and T. Sós and Turán) that

$$f\big(n, K^r(t, \ldots, t)\big) = O(n^{r-1/t^{r-1}}). \tag{25}$$

I deduced from (25) the following result: Let $1 \leqslant b_1 < b_2 < \ldots$ be an infinite sequence of integers. Denote by $g(n)$ the number of solutions of $n = b_i b_j$. Assume that $g(n) > 0$ for all n. Then

$$\limsup_{n=\infty} g(n) = \infty. \tag{26}$$

The additive analog of (26) is an old problem of Turán and mine, and, as stated in Chapter 1, is still open [24].

I just state one more theorem of Simonovits and mine which has direct consequences to some of the problems discussed in Chapter 2. We proved that (for $n > n_0$)

$$f\big(n; K(r, r, 1)\big) = \left[\frac{n^2}{4}\right] + \frac{n}{2} + 1 \tag{27}$$

and (27) implies that for $n \equiv 0 (\bmod 8)$

$$f_4(n) = \frac{n^2}{4} + \frac{n}{2} + 1. \tag{28}$$

The simplicity of (28) is in curious contrast to the difficulty of (10) [25].

Now I have to say a few words on Ramsey's theorem. Very much work has been done on this subject and to save space I only state one or two of my favourite problems and refer to the literature. (The list of references is, of course, far from complete [26].) Let $G_1, \ldots, G_k$ be k graphs, and let $r(G_1, \ldots, G_k)$ be the smallest integer n for which if one colours the edges of the complete graph $K(n)$ by k colours arbitrarily, then for some i, $1 \leqslant i \leqslant k$ the i-th colour contains G_i as a subgraph.

It is surprisingly difficult to get *good* upper or lower bounds for these functions, e.g. it is not yet known whether the limit of $r\big(K(m); K(m)\big)^{1/m}$ exists. It is known that it is between $2^{1/2}$ and 4. The sharpest known inequality for $r\big(K(3), K(m)\big)$ states

$$\frac{c_2 m^2}{(\log m)^2} < r\big((K(3), K(m)\big) < \frac{c_1 m^2}{\log m}. \tag{29}$$

The proof of (29) uses probabilistic methods. Presumably

$$r\big(K(r), K(m)\big) > \frac{cm^{r-1}}{(\log m)^{a_r}} \tag{30}$$

for some constant a_r, but (30) resisted so far all attempts for $r > 3$. It seems very likely that

$$r\big(K(m), C_4\big) < m^{2-\varepsilon} \tag{31}$$

holds, but it is not even known that (C_3 is a triangle, $C_3 = K(3)$)

$$r\big(K(m), C_4\big)/r\big(K(m), C_3\big) \to 0. \tag{32}$$

Szemerédi recently observed that

$$r\big(K(m), C_4\big) < \frac{cm^2}{(\log m)^2}. \tag{33}$$

(33), in view of (31), only just fails to prove (32). Ajtai, Komlós and Szemerédi [8] in fact proved the following lemma, which immediately gives (33), and was crucial to the proof of (7):

Trivially, every $G(n; kn)$ has an independent set of size $> n/2k$. Now, if one assumes that our $G(n; kn)$ has no triangles, then the largest independent set has size $> (cn \log k)/k$ (which, apart from the value of the constant, is best possible). In fact, the result remains true even if we assume only that the number of triangles is abnormally small. Several unsolved problems remain, e.g. if we assume only that our $G(n; kn)$ contains no $K(r)$, can we ensure an independent set of size much larger than $n/2k$. The results in this case are not yet in their final form [27].

I just mention one application of Ramsey's theorem. 50 years ago E. Klein asked: Is there an $f(n)$ such that if $x_1, \ldots, x_{f(n)}$ is a set of $f(n)$ points in the plane, no three on a line, then one can always find a subset of n points forming the vertices of a convex n-gon. Szekeres deduced this from Ramsey's theorem. He also conjectured that $f(n) = 2^{n-1}+1$. Later we proved

$$2^{n-1}+1 \leqslant f(n) \leqslant \binom{2n-4}{n-2}.$$

The first unsettled case is whether $f(6) = 17$ or not.

To finish the paper I want to state a conjecture of mine which would have some geometric applications: Is it true that there is an $n = n(\varepsilon)$ so that if $\varepsilon n < k < (\frac{1}{2}-\varepsilon)n$ and $|S| = n$, $A_i \subset S$, $1 \leqslant i \leqslant T_n$ is a family of subsets of S so that for every $1 \leqslant i_1 < i_2 \leqslant T_n$

$$|A_{i_1} \cap A_{i_2}| \neq k,$$

then $T_n < (2-\varepsilon)^n$. Peter Frankl just proved this conjecture.

References

[1] Here is a partial list of my papers on these subjects:

I. Extremal Problems in Number Theory. In: A. L. Whiteman (ed.), *Proc. Symp. Amer. Math.* VIII, *Theory of Numbers*, Amer. Math. Soc., 1963, pp. 181–189.

II. Combinatorial Problems in Geometry and Number Theory. In: D. K. Ray-Chaudhuri (ed.), *Ibid.* XXXIV, *Relations between Combinatorics and Other Parts of Mathematics*, Amer. Math. Soc., 1979, pp. 149–162.

III. On the Applications of Combinatorial Analysis to Number Theory, Geometry and Analysis, *Actes du Congres Int. des Math. Nice 3* (1970), pp. 201–210.

IV. Remarks on Number Theory IV and V, Extremal Problems in Number Theory (in Hungarian), *Mat. Lapok* **13** (1962), pp. 228–253, and *ibid.* **17** (1966), pp. 135–155.

V. On Some Problems in Elementary and Combinatorial Geometry, *Annali di Mat. Pura et Applicata* **53** (1975), pp. 99–108. Most of the results in the chapter on geometry for which I do not give references are found in this paper.

VI. On the Combinatorial Problems which I would Most Like to See Solved, *Combinatorica* **1** (1981), pp. 25–42.

VII. Some Old and New Problems in Various Branches of Combinatorics, *Proc. Tenth Southeastern Conference in Combinatorics Congressus Num. XXIII*, 1979, Vol. 1, pp. 19–37. This paper lists many of my relevant papers. For a large list of problems in combinatorial and elementary number theory see P. Erdös and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, *Monographie de l'Enseignement Math.* **28**, Geneve, 1981, and R. K. Guy, Usolved Problems in Number Theory, *Problem Books in Math.* 1, edited by P. R. Halmos, Springer Verlag, 1982. For problems in geometry see the booklet of W. Moser, *Research Problems in Discrete Geometry*, Dept. of Math., McGill Univ., 1981.

[2] Graham R. L., Rothschild B. L., and Spencer J. H., *Ramsey Theory*, A. Wiley, Interscience Publication, 1980.

[3] Fürstenberg H., *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[4] Tchudakoff N. G., Theory of the Characters of Number Semigroups, *Journal Indian Math. Soc.* **20** (1951), pp. 11–15.

[5] Spencer J., Problem 185, *Canad. Math. Bull.* **16** (1973), p. 464.

[6] Halberstam H. and Roth K. F., *Sequences*, Oxford Univ. Press, 1960. Recently this book was reprinted by Springer-Verlag.

[7] Guy R. K., Sets of Integers, whose Subsets Have Distinct Sums. In: A. Rosa, E. Salidumi and J. Turgeon (eds.), *Theory and Practice of Combinatorics*, Annals of Discrete Math., dedicated to A. Kotzig, North-Holland, 1982, pp. 141–154.

[8] Ajtai M., Komlós J., and Szemerédi E., A Dense Infinite Sidon Sequence, *European J. of Combinatorics* **2** (1981), pp. 1–11.

[9] Komlós J., Pintz J., and Szemerédi E., On Heilbronn's Triangle Problem, *J. London Math. Soc.* **25** (1982), pp. 13–29.

[10] Erdös P., Some Applications of Ramsey's Theorem to Additive Number Theory, *Europ. J. of Combinatorics* **1** (1980), pp. 43–46.

[11] Komlós J., Sulyok M., and Szemerédi E., Linear Problems in Combinatorial Number Theory, *Acta Math. Acad. Sci. Hungar.* **26** (1975), pp. 113–121.

[12] Graham R. L. and Sloane M. T. A., S. I. A. M. J., *Algebraic and Discrete Methods*, On Additive Bases and Harmonious Graphs, Vol. 1 (1980), pp. 382–404.

[13] Lagarias J., Odlyzko A. M., and Shearer J., On the Density of Sequences of Integers the Sum of No Two of Which Is a Square, I: Arithmetic Progressions, II: General Sequences, *J. Comb. Theory, Ser. A* **33** (1982), pp. 167–184 and **34** (1983), pp. 123–139.

[14] Burr S. A., Grünbaum B., and Sloane N. J. A., The Orchard Problem, *Geometriae Dedicata* **2** (1974), pp. 397–424. This paper has extensive references and many open problems and interesting historical remarks.

[15] *Proc. Symp. in Pure Math. VII, Convexity*, Amer. Math. Soc., 1963. In particular: Grünbaum B., Borsuk's Problem and Related Questions, pp. 271–284, for further problems see also Danzer L., Grünbaum B., and Klee V., Helly's Theorem and Its Relatives, pp. 101–180.

For further problems and results on combinatorial geometry see e. g. Grünbaum B., Arrangements and Spreads, *Regional conference series in Math.* **10**, Amer. Math. Soc. Providence (1972); Motzkin Th., The Lines and Planes Connecting the Points of a Finite Set, *Trans. Amer. Math. Soc.* **70** (1951), pp. 451–464; Hadwiger, Debrunner and Klee, *Combinatorial Geometry in the Plane*, Holt, Rinehart and Winston.

Erdös P., Graham R. L., Montgomery P., Rothschild R. L., Spencer T., and Straus E. G., Euclidean Ramsey Theorems, I, II and III, *Journal Comb. Theory, Ser. A* **14** (1973), pp. 341–363; *Infinite and Finite Sets*, Coll. Math. J. Bolyai, North-Holland, 1973, pp. 529–557 and 559–583.

Grünbaum B., *Arrangements and Spreads*, Amer. Math. Soc. (1972).

[16] Burr S. A., Generalized Ramsey Theory for Graphs – A Survey. In: *Graphs and Combinatorics*, Springer-Verlag, Berlin, 1974, pp. 53–75; also: A Survey of Noncomplete Ramsey Theory for Graphs, *Annals of the New York Academy of Sciences, Topics in Graph Theory* **328** (1979), pp. 58–75.

Chung F. R. K., and Grinstead L. M., A Survey of Bounds for Classical Ramsey Numbers, *Journal of Graph Theory* **7** (1983), pp. 25–38.

Erdös P., and Kleitman D. J., Extremal Problems among Subsets of a Set, *Proc. Second Chapel Hill Conf. on Comb. Math.*, Univ. of North Carolina, Chapel Hill, 1970, pp. 146–170; see also *Discrete Math.* **8** (1974), pp. 281–294, and Erdös P., *Art of Counting*, M.I.T. Press, 1973.

Erdös P. and Simonovits M., Some Extremal Problems in Graph Theory, Combinatorial Theory and Its Application, *Proc. Coll. Balatonfüred* (1969), pp. 377–390.

Katona Gy., Extremal Problems for Hypergraphs, *Combinatorics, Proc. NATO Advanced Study, Breukelen 1979*, Part 2, pp. 13–42.

Kleitman D. J., On an Extremal Property of Antichains in Partial Orders, *ibid.* pp. 77–90.

[17] Bollobás B., *Extremal Graph Theory*, Academic Press, London and New York, 1978.

Simonovits M., Extremal Graph Theory. In: Beineke and Wilson (eds.), *Selected Topics in Graph Theory II*, Academic Press.

Simonovits M., Extremal Graph Problems, Degenerate Extremal Problems and Supersaturated Graphs, *Proc. Conference on Silver Jubilee of University of Waterloo* (1982).

[18] Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Zeitschrift* **27** (1928), pp. 544–548.

[19] Behrend F., On Sequences of Numbers Not Divisible One by Another, *J. London Math. Soc.* **10** (1935), pp. 42–49; see laso: Elliott T. D. A., *Probabilistic Number Theory*, Vol. 1, *Grundlehren der Math. Wissenschaften* **239**, Springer-Verlag, pp. 214–216.

[20] Erdös P., Sàrközy A., and Szemerédi E., On Divisibility Properties of Sequences of Integers, *Number Theory, Coll. Math. Soc. J. Bolyai* **2**, North-Holland, 1970, pp. 35–47.

[21] Berge C., *Graphs et hypergraphs*, Monographies Universitaires de Mathematiques, Dunod, 1970.

[22] Kostochka A. V., A Class of Constructions for Turán's 3-4 Problem, *Combinatorica* **2** (1982), pp. 192–198.

[23] Erdös P., On some Applications of Graph Theory to Number Theoretic Problems, *Publ. Ramanujan Inst.* **1** (1969), pp. 131–136.

[24] Erdös P., On Extremal Problems of Graphs and Generalized Graphs, *Israel T. Math.* **2** (1964), pp. 183–190; and On the Multiplicative Representation of Integers, *ibid.*, pp. 251–261. For further problems on hypergraphs see Erdös P., On Some Extremal Problems on r-Graphs, *Discrete Math.* **1** (1971), pp. 1–6.

[25] Erdös P., On Some Applications of Graph Theory to Geometry, *Canad, J. Math.* **19** (1967), pp. 968–971; Simonovits M., A Method for Solving Extremal Problems in Graph Theory Stability Problems, *Proc. Coll. on Graph Theory Held in Tihany (Hungary) 1966*; see also: Erdös P., Extremal Problems in Graph Theory, *Proc. Symp. Graph Theory, Smolenice 1963*, pp. 29–36. For further extremal results see Erdös P. and Simonovits M., An Extremal Graph Problem, *Acta Math. Acad. Sci. Hung.* **22** (1971), pp. 275–282.

[26] See e.g. the papers of Burr quoted in [16]; also: Erdös P. and Spencer J., *Probabilistic Methods in Combinatorics*, Acad. Press, New York, 1974.

[27] Ajtai M., Erdös P., Komlós J., and Szemerèdi E., On Turàn's Theorem for Sparse Graphs, *Combinatorica* **1** (1981), pp. 313–318.

[28] Erdös P. and Szekeres E., A Combinatorial Problem in Geometry, *Compositio Math.* **2** (1935), pp. 963–970; and: On Some Extremal Problems in Elementary Geometry, *Ann. Univ. Sci. Budapest. Sect. Math.* **3–4** (1961), pp. 53–62; see also: *The Art of Counting, Selected Writings of P. Erdös*, MIT Press, 1973.

[29] Frankl P. and Wilson R. M., Intersection Theorems with Geometric Consequences, *Combinatorica* **1** (1981), pp. 357–368.

[30] Sós V. T., Irregularities of Partitions, Ramsey Theory, Uniform Distribution, *Surveys in Combinatorics*, Invited papers for the 9th British Combinatorial Conference 1983, pp. 201–245.

[31] Ungàr P., $2n$ Non Collinear Points Determine at least $2n$ Directions, *J. Comb. Theory* **33** (1982), pp. 343–347.

[32] Beck J., On a Problem of K. F. Roth Concerning Irregularities of Point Distributions.

[33] Pritchard P.A., Eighteen Primes in an Arithmetic Progression, *Math. Comp.* **41** (1983), p. 697.

Proceedings of the International Congress of Mathematicians
August 16-24, 1983, Warszawa

WENDELL H. FLEMING

Optimal Control of Markov Processes*

1. Introduction

The purpose of this article is to give an overview of some recent developments in optimal stochastic control theory. The field has expanded a great deal during the last 20 years. It is not possible in this overview to go deeply into any topic, and a number of interesting topics have been omitted entirely. The list of references includes several books, conference proceedings and survey articles.

The development of stochastic control theory has depended on parallel advances in the theory of stochastic processes and on certain topics in partial differential equations. On the probabilistic side one can mention decomposition and representation theorems for semimartingales, formulas for absolutely continuous change of probability measure (e.g. the Girsanov formula), and the study of Ito-sense stochastic differential equations with discontinuous coefficients. It seems fair to say that these developments in stochastic processes were in turn to an extent influenced by their applications in stochastic control. For controlled Markov diffusion processes, there is a direct connection with certain nonlinear partial differential equations via the dynamic programming equation. These equations are of second order, elliptic or parabolic, and possibly degenerate. Stochastic control gives a way to represent their solutions probabilistically. There is an unforeseen connection with differential geometry via the Monge–Ampère equation.

Broadly speaking, stochastic control theory deals with models of systems whose evolution is affected both by certain random influences

* This research has been supported in part by the Air Force Office of Scientific Research under contract #AFOSR-81-0116 and in part by the National Science Foundation under contract #MCS-8121940.

and also by certain inputs chosen by a "controller". We are concerned here only with state-space formulations of control problems in continuous time. Moreover, we consider only markovian control problems in which the state x_t of the process being controlled is Markov provided the controller follows a Markov control policy. We shall not discuss at all the extensive engineering literature on input-output formulations particularly for linear system models, see Åström [1].

We shall mainly discuss the case of continuously acting control, in which at each time t a control u_t is applied to the system. However, in § 8 we briefly mention impulsive control problems, in which control is applied only at discrete time instants. In optimal stochastic control theory the goal is to minimize (or maximize) some criterion depending on the states x_t and controls u_t during some finite or infinite time interval. In § 2 we formulate a class of optimal control problems for Markov processes, with criterion (2.2) to be minimized. The distinction between problems in which x_t is known to the controller, and problems with partial observations is made there. When x_t is known, the dynamic programming method can be used. In principle, this method leads directly to an optimal Markov control policy, although it rarely gives the optimal policy explicitly. In § 3, both analytical and probabilistic approaches are indicated. Associated with dynamic programming is the Nisio nonlinear semigroup (§ 4). In § 5 we discuss methods of approximate solution and special problems. In § 6 a logarithmic transformation is applied to positive solutions of the backward equation of a Markov process. There results a controlled Markov process, leading to connections between stochastic control and such topics as stochastic mechanics, large deviations and nonlinear filtering. The case of controlled, partially observed processes is mentioned in § 7, along with adaptive control of Markov processes. Finally in § 9 we indicate a few of the various difficulties encountered in seeking to implement in engineering applications the mathematically sophisticated results of the theory, and mention some newer areas of application.

2. Controlled Markov processes

We consider optimal stochastic control problems of the following kind. We are given metric spaces Σ, U called the *state space* and *control space*, respectively. For each fixed $u \in U$ there is a linear operator L^u which generates a Markov, Feller process with state space Σ. The domain of L^u contains, for each $u \in U$, a set D dense in the space $C(\Sigma)$ of bounded uniformly continuous functions on Σ. The state and control processes

x_t, u_t are defined on some probability space $(\Omega, \mathscr{F}, P)$. The Σ-valued process x_t is adapted to some increasing family of σ-algebras $\mathscr{F}_t \subset \mathscr{F}$, and the trajectories x are right continuous. The U-valued process u_t is predictable with respect to an increasing family of σ-algebras $\mathscr{G}_t \subset \mathscr{F}_t$. The σ-algebra $\mathscr{G}_t$ describes in a measure theoretic way the information available to the controller at time t. The processes (x_t, u_t) are related by the requirement that

$$Mg(t) = g(x_t) - g(x_0) - \int_0^t L^{u_s} g(x_s)\, ds \tag{2.1}$$

is a $(\mathscr{F}_t, P)$ martingale for every $g \in D$. We consider a fixed, finite time interval $0 \leqslant t \leqslant T$, and the objective to minimize a criterion of the form of an expectation

$$J = E\left\{\int_0^T k(x_t, u_t)\, dt + G(x_T)\right\}. \tag{2.2}$$

Example 1. Controlled finite-state Markov chain, with $\Sigma = \{1, 2, \ldots, N\}$. In this case L^u is identified with the infinitesimal matrix (q_{ij}^u) of the chain. When the control u_t is applied, the jumping rate of x_t from state i to j is $q_{ij}^{u_t}$.

Example 2. Controlled diffusion process with $\Sigma = \boldsymbol{R}^n$,

$$x_t = x_0 + \int_0^t f(x_s, u_s)\, ds + \int_0^t \sigma(x_s, u_s)\, dw_s, \tag{2.3}$$

with w_t a brownian motion (of some dimension d) independent of the initial state x_0. In this case

$$L^u = \tfrac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x, u) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} f_i(x, u) \frac{\partial}{\partial x_i} \tag{2.4}$$

with $a = \sigma\sigma'$ and $D = \{g\colon g, g_{x_i}, g_{x_i x_j} \in C(\boldsymbol{R}^n),\ i, j = 1, \ldots, n\}$. The diffusion is called *nondegenerate* if the eigenvalues of $a(x, u)$ are bounded below by $c > 0$.

Further assumptions, which vary from author to author in the literature, need to be made. To avoid undue complication, in the discussion to follow we take a compact control space U, and $k(x, u)$, $G(x)$ bounded, uniformly continuous. In (2.3), $f(x, u)$, $\sigma(x, u)$ are bounded and as smooth as necessary. The σ-algebras $\mathscr{F}_t$, $\mathscr{G}_t$ are right continuous and completed.

If x_t is $\mathscr{G}_t$-measurable, then the controller can observe the state x_t. In this case, one may as well take $\mathscr{G}_t = \mathscr{F}_t$ and known initial state x_0. This is the situation in Sections 3–6 to follow. If (2.1) holds, we call

$$\mathfrak{a} = (\Omega, \mathscr{F}, P, \{\mathscr{F}_t\}, x_., u_.)$$

an *admissible system* for the control problem with completely observed states.

A *Markov control policy* is a Borel measurable function from $[0, T] \times \Sigma$ into U. An admissible system $\mathfrak{a}$ is obtained via a Markov control policy $\underline{u}$ if

$$u_t = \underline{u}(t, x_{t-}). \tag{2.5}$$

Given $\underline{u}$ and $x_0 \in \Sigma$, one would like to know whether a corresponding admissible system exists, with x_t a Markov process. Under sufficiently strong restrictions this is well known. For instance, in case of controlled diffusions a Lipschitz condition on $\underline{u}(t, x)$ would imply the classical Ito conditions. For nondegenerate controlled diffusions, existence follows from Krylov [8, p. 87] for any bounded $\underline{u}$. The Markov property of x_t can be obtained under stronger hypotheses. For instance, for nondegenerate diffusions it holds if in (2.3) $\sigma = \sigma(x)$. A martingale method for obtaining the Markov property is to show that the probability distribution $P^{\underline{u}}_{x_0}$ of the state trajectory $x_.$ is unique and depends continuously on the initial state x_0 [59]. In general x_t is only a weak-sense solution to (2.3), since neither the probability space nor the brownian motion w_t are given in advance. However, in the nondegenerate case with $\sigma = \sigma(x)$ a result of Veretennikov [61] gives a strong solution.

3. Dynamic programming

The dynamic programming approach to the control problem with completely observed states x_t can be described in a purely formal way, as follows. For initial state $x_0 \in \Sigma$ and admissible system $\mathfrak{a}$, write $J = J(T, x_0, \mathfrak{a})$ in (2.2). Let

$$W(T, x_0) = \inf_{\mathfrak{a}} J(T, x_0, \mathfrak{a}). \tag{3.1}$$

Formal reasoning indicates that $W(T, x)$ should satisfy the dynamic programming equation

$$\frac{\partial W}{\partial T} = AW, \quad T > 0, \tag{3.2}$$

with initial data $W(0, x) = G(x)$, where

$$Ag(x) = \min_{u \in U} [L^u g(x) + k(x, u)]. \tag{3.3}$$

Formally, an optimal Markov policy $\underline{u}^*$ is found by requiring $\underline{u}^*(t, x)$ to minimize $L^u W(T-t, x) + k(x, u)$ among all $u \in U$. Instead of the finite time control problem, control until x_t exits a given open set $\mathcal{O} \subset \Sigma$ can be considered. In that case the dynamic programming equation becomes the autonomous form of (3.2) in $\mathcal{O}$, with $W(x) = G(x)$ for $x \in \partial\mathcal{O}$. There are also autonomous dynamic programming equations associated with the infinite time control problem, with discounted cost or average cost per unit time criteria to be minimized.

In the rigorous mathematical treatment of dynamic programming there is one easy result, the so-called *Verification Theorem* [7, p. 159]. Roughly speaking, it states that if $W(T, x)$ satisfying (3.2) with the initial data and the associated Markov policy $\underline{u}^*$ are both "sufficiently regular", then $\underline{u}^*$ is indeed optimal and $W(T, x)$ is the minimum performance in (3.1). The Verification Theorem is used to obtain explicit solutions, in those cases where such a solution is known. Much more difficult are the questions of existence of sufficiently regular W and $\underline{u}^*$, and there is a large literature dealing with various aspects of them. One approach is analytical with the stochastic interpretation made afterward. In this approach, existence of solutions to the dynamic programming equation and their regularity properties are studied, using non-probabilistic methods. It is then proved that optimal (or at least ε-optimal) Markov control policies exist. A second approach is probabilistic. In this approach, one starts with the minimum cost function W in (3.1) and develops stochastic counterparts to the dynamic programming conditions for a minimum. A third approach is to consider an associated nonlinear semigroup (§ 4). While this approach leads to fewer technical difficulties than either of the other two, it also leads to weaker results.

For controlled diffusions the analytical approach is remarkably well developed (see Krylov [8], Lions [45]). In the nondegenerate case the dynamic programming equation is a second order nonlinear partial differential equation of parabolic type also called a Hamilton–Jacobi–Bellman equation. In various other formulations, with x_t controlled for all time $t \geqslant 0$ or until exit from an open set $\mathcal{O}$, the Hamilton–Jacobi–Bellman equation is elliptic rather than parabolic. Under reasonable assumptions the problem, the solution S has generalized second derivatives which are locally bounded. In the elliptic case a deeper regularity result of Evans ([26], [60]) gives a classical solution. In the degenerate case W is less regular with locally bounded generalized first derivatives W_{x_i}. The dynamic programming equation (3.2), suitably interpreted in terms of Schwartz distributions, still holds ([8], [45]). For the case of controlled jump Markov processes, results on existence, uniqueness and regularity of solutions to (3.2) were obtained by Pragarauskas [52].

A large class of nonlinear elliptic or parabolic equations, satisfying appropriate convexity conditions, can be represented as Hamilton–Jacobi–Bellman equations. As Gaveau [35] pointed out, the Monge–Ampère equation has such a representation.

In the probabilistic approach, the starting point is to rewrite the dynamic programming principle in the following martingale form. Given an admissible system α let

$$m_t = \int_0^t k(x_s, u_s)\,ds + W(T-t, x_t).$$

Then m_t is a $(\mathscr{F}_t, P)$ submartingale, and α is optimal if and only if m_t is a $(\mathscr{F}_t, P)$ martingale. With the aid of the Doob–Meyer decomposition for submartingales and some martingale representation theorems, conditions for optimality are obtained (see Bismut [21], Davis [16], Elliott [25], El Karoui [5]). These conditions are probabilistic counterparts of those expressed analytically by the dynamic programming equation (3.2). With the probabilistic approach difficult questions of regularity of solutions to (3.2) are avoided. The probabilistic techniques give results about existence of optimal Markov policies ([21], [5, p. 218]). These methods also give conditions for a minimum for optimal control under partial observations.

A different kind of Markovian control problem for diffusions, in which the control acts only on the boundary of a region $\mathcal{O} \subset \boldsymbol{R}^n$ was considered by Vermes [62].

4. The Nisio nonlinear semigroup

The dynamic programming principle can be restated in another form, in terms of a semigroup of nonlinear operators. In purely formal way, this is done as follows. In (2.2) we fix k but consider various G. We rewrite the infimum in (3.1) as $W(T, x) = S_T G(x)$. The dynamic programming principle is formally equivalent to the semigroup property

$$S_{T_1+T_2} = S_{T_1} \circ S_{T_2} \tag{4.1}$$

of the family $\{S_T\}$ of nonlinear operators. In addition, for "sufficiently regular" G, one should have

$$\frac{d}{dT} S_T G|_{T=0} = AG. \tag{4.2}$$

This formal procedure was put on a rigorous basis by Nisio [10], who obtained $\{S_T\}$ as a semigroup on the space $C(\Sigma)$ and showed under some mild additional conditions that (4.2) holds for $G \in D$ (notation of § 2). Equations (4.1), (4.2) would imply the dynamic programming equation (3.2) if we knew that $W(T, \cdot) = S_T G$ is sufficiently regular (in particular, if S_T maps D into D). However, W does not generally have the desired regularity. In such instances (4.2) is a kind of weaker substitute for (3.2).

Nisio's treatment is analytical. She obtains S_T as the lower envelope of the family of linear semigroups S_T^u, where for constant control $u \in U$ the generator of S_T^u coincides on D with the operator $L^u + k(\cdot, u)$. A stochastic treatment of the Nisio semigroup is given in Bensoussan and Lions [2], and a uniqueness result in case of nondegenerate diffusions in Nisio [51]. El Karoui, Lepeltier, and Marchal [24] used another procedure, and obtained a nonlinear semigroup on a larger space of bounded functions G which are measurable in a suitable sense.

5. Explicit and approximate solutions

In a few instances the dynamic programming equation (3.2) can be solved explicitly. Examples are the well known stochastic linear regulator and Merton's optimal portfolio selection problem [7, pp. 160, 165]. For other special problems the solution can be reduced to a free boundary problem. The boundaries to be determined separate regions where some control constraint holds or not. See for example Karatzas and Benes [40].

When a solution cannot be found by special methods, one can seek an approximate solution to (3.2). One class of approximate methods involve discretizations of (3.2). Among such methods the algorithm of Kushner [9] has a natural stochastic control interpretation. The difference equations associated with the algorithm correspond to the dynamic programming equation for an approximating controlled Markov chain.

For the special case of controlled one-dimensional diffusions, Borkar and Varaiya [22] used a procedure in which piece-wise constant approximating Markov control policies are allowed.

Other results give approximate solutions to (3.2) when the state process x_t is a nearly-deterministic controlled diffusion. In (2.3) let $\sigma = \varepsilon^{1/2}\tilde{\sigma}$. The solution is sought in the form of an asymptotic series in ε. In [29] this is done by expanding the solution $W^\varepsilon(T, x)$ in an asymptotic series. The expansion is valid in regions where the solution $W^0(T, x)$ of the corresponding Hamilton–Jacobi equation is smooth. In [20] Bensoussan obtains an asymptotic expansion, using a stochastic maximum principle instead of (3.2).

6. A logarithmic transformation

Consider a linear operator of the form $L+V(x)$, where L is the generator of a Markov process ξ_t with state space Σ. The initial value problem

$$\frac{d\varphi}{dT} = L\varphi + V(x)\varphi \tag{6.1}$$

with data $\varphi(0, x) = \Phi(x)$ has a probabilistic solution by a well known formula of Feynman–Kac type. For positive solutions of (6.1) another probabilistic representation for $\varphi(T, x)$ can often be found in the following way. The logarithmic transformation $I = -\log\varphi$ changes (6.1) into the nonlinear equation

$$\frac{dI}{dT} = H(I) - V(x), \tag{6.2}$$

$$H(I) = -e^I L(e^{-I}). \tag{6.3}$$

If one can find a control problem of the kind in § 2 such that

$$H(I) = \min_{u\in U}[L^u I + k(x, u)], \tag{6.4}$$

then (6.2) is the dynamic programming equation (3.2). The stochastic control interpretation of $I(T, x)$ is as the minimum of the criterion J in (2.2). Thus, in (3.1) we have $W = I$. For a nondegenerate diffusion obeying the stochastic differential equation

$$d\xi_t = b(\xi_t)dt + \sigma(\xi_t)dw_t, \tag{6.5}$$

a Markov control policy $\underline{u}(t, x)$ changes the generator L to $L^{\underline{u}}$, corresponding to change of drift from $b(x)$ to $\underline{u}(t, x)$ in (6.5). In (2.2) one takes

$$k(x, u) = \tfrac{1}{2}\big(b(x)-u\big)'a^{-1}(x)\big(b(x)-u\big),$$

with $a = \sigma\sigma'$. An appropriate control problem for the case of ξ_t a jump Markov process is described in [31], and for a general class of Markov ξ_t in Sheu's thesis [58]. The change of generator from L to $L^{\underline{u}}$ corresponds to a change of probability measure. It was pointed out by M. Day that this change of measure results by conditioning with respect to $\Phi(x_T)$ (see [31, (4.5)]).

In case $L = \frac{1}{2}\Delta$, corresponding to ξ_t a brownian motion (6.1) is the heat equation with a potential term. The stochastic control interpretation of $S = -\log\varphi$ is as least average action. Upon rescaling, taking $L = \frac{1}{2}\lambda\Delta$ and replacing V by $\lambda^{-1}V$, the usual least action is obtained as a "classical mechanical limit" as $\lambda\to 0$ [28]. The heat equation with potential is the "imaginary time" analogue of the Schrödinger equation of quantum mechanics. There is an intriguing connection between stochastic control and the Schrödinger equation, whose implications are not as yet well understood [36]. This work is in the framework of Nelson's stochastic mechanics. An apparently different theory of "stochastic mechanics" was developed by Bismut [4].

Holland [39] gave a stochastic control interpretation of the dominant eigenvalue of the Schrödinger equation as minimum mean total energy of a particle in equilibrium. The approach was again based on a logarithmic transformation and subsequently led to Sheu's treatment [58] of the Donsker–Varadhan formula for the dominant eigenvalue of the operator $L+V$ appearing in (6.1).

The Ventsel–Freidlin theory of large deviations deals with asymptotic probabilities of rare events associated with nearly deterministic Markov processes. The logarithmic transform gives another approach to results of this kind. As an illustration we consider the problem of exit from an open set $D\subset\Sigma$ during the time interval $0\leqslant t\leqslant T$. Let x_t^ε be a Markov process tending to a deterministic limit x_t^0 as $\varepsilon\to 0$. Let $I^\varepsilon = -\varepsilon\log P_x(\tau^\varepsilon\leqslant T)$, where τ^ε is the exit time of x_t^ε from D. Under various assumptions (including a suitable scaling of ε), I^ε tends to a limit I^0, where $I^0(T, x)$ is the minimum of a certain "action functional" among curves starting at $x\in D$ and leaving D by time T. In the stochastic control approach $I^\varepsilon(T, x)$ is the minimum performance in a corresponding stochastic control problem [27], [31], [58]. In this approach a minimum principle is associated with the large deviation problem for $\varepsilon > 0$, not just in the limit as $\varepsilon \to 0$.

In [32], the logarithmic transformation was applied to solutions to the pathwise equation of nonlinear filtering, making a connection between filtering and stochastic control.

7. Partial observations; adaptive control

The states x_t of a stochastic system often cannot in practice be measured directly, or perhaps can only be measured with random errors. This has led to an extensive literature or nonlinear filtering and on optimal control under partial observations. For controlled diffusions, a standard model is to take state dynamics (2.3) and an observation process y_t governed by

$$y_t = \int_0^t h(x_s)\,ds + W_t, \tag{7.1}$$

with W a brownian motion independent of w. The information available to the controller at time t is usually assumed to be described by the σ-algebra $\mathscr{G}_t$ generated by observations y_s for $s \leqslant t$. However, existence of optimal controls has been proved only with a somewhat wider class of admissible controls than those adapted to this family $\{\mathscr{G}_t\}$.

Several good survey articles on controlled partially observed diffusions have recently appeared [15], [16], [17]. Hence, we shall not try to summarize the various results here. In studying partially observed control problems it is useful to introduce an auxiliary "separated" control problem. In the separated problem the role of "state" process is taken by a measure-valued stochastic process σ_t [34]. The measure σ_t represents an unnormalized conditional distribution of x_t given observations and controls $y_s, u_s, 0 \leqslant s \leqslant t$. A nonlinear semigroup for the controlled, measure-valued process σ_t has been constructed [19], [30], [33]. Among other recent work, we mention that of Rishel [53] on partially observed jump processes, and of Mazziotto and Szpirglas [49] on impulsive control under partial information.

In *adaptive control* the objective is the simultaneous control and identification of unknown system parameters. Common techniques in discrete-time adaptive control involve sequential techniques, based on maximum likelihood or least squares, for updating estimates of unknown parameters. In the context of adaptive control of Markov chains see the pioneering work of Mandl [48], also Borkar and Varaiya [23], Kumar and Lin [41]. Another (Bayesian) viewpoint is to treat adaptive control of Markov processes as a special case of stochastic control under partial observations. This is done by simply regarding the unknown parameters as additional (nontime-varying) components of the system state. From a practical standpoint this approach encounters well known difficulties, in that effective solutions to partially observed stochastic problems are difficult to obtain. Nevertheless, special cases in which the problem becomes finite-dimensional have been treated by Hijab [38] and Rishel [54].

8. Impulse control; problems with switching costs

In impulse control problems the control actions are taken at discrete (random) time instants, and each control action leads to an instantaneous change in the state x_t. Typical impulse control problems are those of stock inventory management, in which a control action is to reorder with immediate delivery of the order.

The analytic treatment of impulse control was initiated and developed systematically by Bensoussan and Lions [3], with emphasis on the control of nondegenerate diffusions. The dynamic programming equation is replaced by a set of inequalities which take the form of a quasivariational inequality. For the case of degenerate diffusions see Menaldi [50], and for impulsive control for Markov–Feller process see Robin [55], [56]. Lepeltier and Marchal [43] gave a probabilistic treatment.

Another class of stochastic control problems of recent interest are those in which control actions are taken at discrete time instants, with no instantaneous change in x_t but with a cost of switching control actions. Such problems arise in the theory of controlled queues (see Sheng [57]) and in control of energy generating systems under uncertain demand. The analytical treatment again is to reduce the problem to a quasi-variational inequality (see Lenhart and Belbas [42], Liao [44]).

9. Applications

Optimal stochastic control theory was initially motivated by problems of control of physical devices. More recent influences have come from management science, economics, and information systems. Until now, the impact on engineering practice of much of the sophisticated mathematical theory has been small. The stochastic linear regulator is a standard tool, because the optimal Markov control policies turn out to be linear in the state x. If the Markov policy is nonlinear, it is difficult to implement. Moreover, other issues may be considered in practice more important than optimality of system performance as predicted by the stochastic control model. The model is generally a simplification of nature, through linearizations, reductions of dimensionality, assumptions that noises are white, etc. A control which performs well (even optimally) according to the model may behave poorly in a real control system. The question of robustness of controls with respect to unmodelled system dynamics is of current interest in the engineering control literature (see for example [63]). A different sort of question is that of stochastic controllability [64].

We conclude by mentioning two novel applications of stochastic control. One is Arrow's model of exploration consumption, and pricing of a randomly distributed natural resource. This model was analyzed in detail by Hagan, Caflisch and Keller [37]. They determined approximately the free boundary between portions of the state space where new exploration should or should not be undertaken.

Ludwig and associates have applied a stochastic control method to fishery management problems [47]. The fishery resource is controlled through the rate at which fish are harvested. This work has an important statistical aspect as well as the control aspect, since errors in measuring unknown parameters in the fishery model can be important.

References

Books and Monographs

[1] Åström K., *Introduction to Stochastic Control Theory*, Academic Press, 1970.

[2] Bensoussan A. and Lions J. L., *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris, 1978.

[3] Bensoussan A. and Lions J. L., *Contrôle impulsionnel et inéquations quasi-variationnelles*, Dunod, Paris, 1981.

[4] Bismut J.-M., *Mécanique aléatoire*, Springer Lecture Notes in Math. No. 866, 1981.

[5] El Karoui N., *Les aspects probabilistes du contrôle stochastique*, École d'Été de Probabilités de Saint-Fleur IX 1979, Springer Lecture Notes in Math. No. 876, 1981, pp. 73–238.

[6] Elliott R. J., *Stochastic Calculus and Applications*, Springer-Verlag, 1982.

[7] Fleming W. H. and Rishel R. W., *Deterministic and Stochastic Optimal Control*, Springer-Verlag, 1975.

[8] Krylov N. V., *Controlled Diffusion Processes*, Springer-Verlag, 1980.
[9] Kushner H. J., *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, 1977.
[10] Nisio M., *Lectures on Stochastic Control Theory*, Indian Statistical Institute Lecture Notes, Series No. 9, 1981.

Conference Proceedings

[11] Fleming W. H. and Gorostiza L. G. (eds.), *Advances in Filtering and Optimal Stochastic Control*, Springer Lecture Notes in Control and Inf. Sci. No. 42, 1982.
[12] Kohlmann M. and Christopeit N. (eds.), *Stochastic Differential Systems*, Springer Lecture Notes in Control and Inf. Sci. No. 43, 1982.
[13] Kohlmann M. and Vogel W. (eds.), *Stochastic Control Theory and Stochastic Differential Systems*, Springer Lecture Notes in Control and Inf. Sci. No. 16, 1979.

Survey Articles

[14] Bismut J. M., An Introductory Approach to Duality in Optimal Stochastic control, *SIAM Review* **20** (1978).
[15] Christopeit N. and Kohlmann M., Some Recent Results on the Control of Partially Observable Stochastic Systems, in Ref. [12], pp. 251–275.
[16] Davis M. H. A., Martingale Methods in Stochastic Control, in Ref. [13], pp. 85–117.
[17] Davis M. H. A., Some Current Issues in Stochastic Control Theory, *Stochastics*, to appear.
[18] Lions P.-L., Optimal Control of Diffusion Processes and Hamilton–Jacobi–Bellman Equations, These Proceedings, pp. 1403–1418.

Other Articles

[19] Bensoussan A., in Ref. [11], pp. 38–53.
[20] Bensoussan A., in Ref. [12], pp. 216–228.
[21] Bismut J. M., *Memoirs Amer. Math. Soc.* **167**, 1976.
[22] Borkar V. and Varaiya P., *IEEE Trans. Auto. Control* **AC-26** (1981), pp. 466–470.
[23] Borkar V. and Varaiya P., *SIAM J. Control Optimiz.* **20** (1982), pp. 470–489.
[24] El Karoui N., Lepeltier J. P. and Marchal B., in Ref. [12], pp. 294–301.
[25] Elliott R. J., in Ref. [13], pp. 142–155.
[26] Evans L. C., *Trans. Amer. Math. Soc.* **275** (1983), pp. 245–255.
[27] Fleming W. H., *Applied Math. Optimiz.* **4** (1978), pp. 329–346.
[28] Fleming W. H., *J. Optimiz. Theory Applic.* (1983).
[29] Fleming W. H., *SIAM J. Control* **9** (1971), pp. 473–517.
[30] Fleming W. H., *SIAM J. Control Optimiz.* **20** (1982), pp. 286–301.
[31] Fleming W. H., in Ref. [11], pp. 131–141.
[32] Fleming W. H. and Mitter S., *Stochastics* **8** (1982), pp. 63–77.
[33] Fleming W. H. and Nisio M., submitted to *Nagoya Math. J.*
[34] Fleming W. H. and Pardoux E., *SIAM J. Control Optimiz.* **20** (1982), pp. 261–285.
[35] Gaveau B., *J. Funct. Anal.* **25** (1977), pp. 391–411.
[36] Guerra F. and Morato L. M., *Physical Review* **D15**, Febr. 1983.
[37] Hagan P. S., Caflisch R. E., and Keller J. B., submitted to *Econometrica.*
[38] Hijab O., *IEEE Trans. Auto. Control*, Febr. 1983.

[39] Holland C. J., *Commun. Pure Appl. Math.* **30** (1977), pp. 755–765; **31** (1978), pp. 509–519.
[40] Karatzas I. and Benes V. E., *SIAM J. Control Optimiz.* **19** (1981), pp. 283–332.
[41] Kumar P. R. and Lin W., *IEEE Trans Auto. Control* **AC-27** (1982), pp. 765–774.
[42] Lenhart S. M. and Belbas S. A., *SIAM J. Appl. Math.*, 1983.
[43] Lepeltier J. P. and Marchal B., *Stochastics* **2** (1979), pp. 243–286.
[44] Liao Y-C., Thesis, Brown Univ., 1982.
[45] Lions P-L., in Ref. [11], pp. 199–215.
[46] Lions P-L. and Menaldi J-L., *SIAM J. Control Optimiz.* **20** (1982), pp. 58–81, 82–95.
[47] Ludwig D. and Walters C. L., *Ecological Modelling* **14** (1982), pp. 273–292.
[48] Mandl P., *Adv. Appl. Prob.* **6** (1974), pp. 40–60.
[49] Mazziotto G. and Szpirglas J., *Proc. 3rd IMA Conf. on Control Th.*, Academic Press, 1981, pp. 567–586.
[50] Menaldi J-L., *SIAM J. Control Optimiz.* **18** (1980), pp. 722–739.
[51] Nisio M., in Ref. [11], pp. 267–275.
[52] Pragarauskas H., in Ref. [12], pp. 338–344.
[53] Rishel R. W., *IEEE Trans. Auto. Control* **AC-22** (1977).
[54] Rishel R. W., *IEEE Trans. Auto. Control* **AC-26** (1981), pp. 606–609.
[55] Robin M., *SIAM J. Control Optimiz.* **19** (1981), pp. 333–358.
[56] Robin M., Thèse d'etat, Univ. de Paris IX, 1978.
[57] Sheng D. D., in Ref. [11], pp. 353–375.
[58] Sheu S. J., Thesis, Brown Univ., 1982.
[59] Stroock S. and Varadhan S. R. S., *Multidimensional Diffusion Processes*, Springer-Verlag, 1979.
[60] Trudinger N., *Math. Research Report*, No. 23, 1982, Australian National University.
[61] Veretennikov A. Yu., *Math. USSR Sbornik* **39** (1981), pp. 387–403.
[62] Vermes D., in Ref. [12], pp. 369–377.
[63] Wittenmark B. and Åström K. J., *Proc. 21st IEEE Conf. on Decision and Control*, Dec. 1982, pp. 33–37.
[64] Zabczyk J., in [12], pp. 144–154.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

C. HOOLEY

Some Recent Advances in Analytical Number Theory

The realm of the analytical theory of numbers is nowadays too vast for one to attempt a complete survey within an article of this length. We therefore mainly restrict ourselves to those aspects of the additive theory that are associated with the author's recent work.

The circle method of Hardy and Littlewood plays a dominant rôle in the analytic part of the additive theory of numbers. Familiar though this method is to experts in the field, it is appropriate in an expository article that we should give a brief description of the underlying procedure in order that we should be aware of its limitations and of the relevance to it of recent mathematical developments.

Avoiding complete generality for the sake of brevity and clarity, we can indicate the nature of the method by considering its formal application to the problem of determining whether an indeterminate equation

$$f(l_1, l_2, \ldots, l_r) = 0 \tag{1}$$

is soluble, where $f(x_1, \ldots, x_r)$ is a polynomial with rational integral coefficients. It being inherent in the technique that normally it should only be applied when the answer to the proposed question is thought to be in the affirmative, the method usually not only settles the problem of existence but supplies an estimate for the number $\nu(x)$ of solutions of (1) that lie in some large appropriate region R_x, where x is a parameter tending to infinity. In many, but by no means all, of the more important problems the natural form of this region is inherent in the other data and is therefore not the subject of a special definition; this, for example, is the situation in Waring's problem when we consider the representation of large numbers N as the sum of s k-th non-negative powers.

The genesis of the method, as modified by Vinogradov, is the observation that

$$\nu(x) = \int_0^1 \sum_{(l_1,\ldots,l_r)\in R_x} e^{2\pi i f(l_1,\ldots,l_r)\theta}\,d\theta.$$

To treat the integral the range of integration is split up into intervals (or arcs as they are usually called, since the procedure is easily interpreted in terms of the circumference of the unit circle) that are in some sense centred by rational numbers (Farey fractions) of the form h/k, where

$$(h,k)=1, \quad 0\leqslant h<k, \quad k\leqslant X,$$

and where X is a suitable function of x. When θ is at the "centre" h/k of an arc and k is small, the integrand can be estimated with great accuracy because

$$e^{2\pi i f(l_1,\ldots,l_r)h/k}$$

is a periodic function in $l_1,\ldots,l_r$ with small periods; consequently, by partial summation or some equivalent process, the integrand can also be satisfactorily calculated when θ is close to h/k. Thus part of the integral can be adequately treated, while the form of the calculations suggests that the residual part is negligible in circumstances where we may reasonably expect there to be an asymptotic formula for $\nu(x)$.

To validate the asymptotic formula thus suggested it is requisite to overcome the difficulties encountered when k is large or when θ is far from the centre of the arc it lies in (these two possibilities are partly interchangeable because there is usually some latitude in the choice of X). There are two main lines of development here. The first is for us to refine the calculations already made so that they are applicable to the entire range, endeavouring in some places to gain improvements by shewing there is some cancellation between contributions from arcs corresponding to a common denominator k. However, matters of such nicety intervene that it is seldom that the programme succeeds. This approach was first used by Kloosterman in his investigation on quaternary quadratic forms, whence flows the present custom of designating such a procedure by the term Kloosterman refinement.

The second and more common technique is applicable to problems that are additive or that can be made additive by a suitable transformation. The Waring's problem about the number $\nu(N)$ of solutions of

$$l_1^k + \dots + l_s^k = N$$

typifying the situation to be covered by this routine, the integrand is now

$$f^s(\theta)\, e^{-2\pi i N\theta},$$

where

$$f(\theta) = \sum_{l \leqslant N^{1/k}} e^{2\pi i l^k \theta}. \tag{2}$$

Over the set $\mathscr{M}$ of minor arcs, which the residual range of integration is termed, the integrand is bounded by

$$(\overline{\underset{\theta \in \mathscr{M}}{\mathrm{bd}}}\, |f(\theta)|)^g \int_0^1 |f(\theta)|^{s-g}\, d\theta,$$

where $0 < g < s$. In favourable circumstances, which alas do not too often occur, the integral above can be estimated because it has a natural arithmetical meaning. The upper bound, on the other hand, has often been satisfactorily estimated through the work of Weyl, Weil, and Vinogradov. Impressive developments in Waring's and other problems have been achieved by these means. There is, however, the substantial shortcoming that the method is inapplicable whenever the order of magnitude of $\nu(N)$ is not larger than N. Consequently, it cannot deal with important unsolved problems such as the Goldbach problem or Waring's problem for four cubes. Similar remarks relate to the more general earlier context when $\nu(x)$ is small in terms of x, where as before R_x is chosen in a natural way. Finally, many variations in this line of attack have been introduced by various writers, and we refer the interested reader to the several treatises on the subject for further details.

Enough has been said already to see the potential relevance of exponential sums of the type

$$\sum_{l_1,\dots,l_r \leqslant u} e^{2\pi i f(l_1,\dots,l_r)h/k} \tag{3}$$

to the circle method. The study of such sums can be easily reduced to that of complete sums

$$\sum_{0 < l_1, l_2, \dots, l_r \leqslant k} e^{2\pi i f(l_1,\dots,l_r)h/k}, \tag{4}$$

which themselves are a specialization of sums of the form

$$\sum_{\substack{0<l_1,\ldots,l_r\leqslant k\\ g(l_1,\ldots,l_r)=0}} e^{2\pi i f(l_1,\ldots,l_r)h/k}, \tag{5}$$

where the significance of the notation f may change as we go from (3) to (5). Although the importance of (4) has been long understood, it has been perhaps less appreciated that there are a multitude of ways in which (5) might conceivably be of assistance.

Best possible estimates for (3) in the case $r = 1$ were made available by Weil's work and allowed considerable progress to be made in additive number theory (both by the circle method and by other means). It was therefore to be expected that Deligne's fundamental and far reaching proof of the generalized Weil conjectures should lead to further advances as soon as it could be shewn how his work was applicable to the sums (5). Estimates for special cases having been obtained by various writers, best possible estimates in the general case were in fact first obtained by the speaker in 1979 by a very simple method ([7], [10], and in a paper shortly to appear), while shortly afterwards Katz obtained similar estimates by a more recondite method, which also shed much light on the structure of the L-functions over algebraic varieties [14].

Recently, a striking advance has been made by Heath-Brown using a Kloosterman refinement and the estimates for multiple exponential sums. As the culmination of a series of important papers, Davenport shewed that an n-ary cubic form f with integral coefficients had a non-trivial integral zero provided that $n \geqslant 16$ ([1], [2], [3]). The result is false for $n = 9$ even when f is non-singular but it had been conjectured that it is always true for $n = 10$. Heath-Brown [4] has proved its truth for $n = 10$ when the form is non-singular, an achievement that settles the situation for the most important category of cubic forms.

Notwithstanding the potential relevance of the Deligne estimates to the circle method, no other significant advance has yet been made through this order of ideas. This is due in part to certain deficiencies in the circle method to which we shall later allude and also to the fact that in many of the more important outstanding problems the expected value of $\nu(x)$ is too small for the method to be applied in any but the most abstruse manner.

Yet there is a further possible avenue of advance through the circle method that seems not yet to have been exploited. This is to go beyond the Kloosterman refinement and to consider possible cancellations between contributions from integrals corresponding to different values of k Serious arithmetical and analytic difficulties, not yet normally capable of resolution, lie athwart this path. But the author has been successful in directing this idea to the theory of indefinite and definite ternary quadratic forms, in which the cardinality of the representations of numbers is too small for a Kloosterman refinement to be adequate. Interesting though this development may be from a methodological angle, it enables no real progress to be made since the theory of ternary forms has already been successfully treated by other more appropriate methods.

We turn now to some recent progress in additive number theory that has been made by alternative methods. First, we mention the mixed problem of representing numbers as the sum of squares and non-negative cubes, the history of which goes back to Hardy and Littlewood. Although it is conjectured that all large numbers are both the sum of one square and two cubes and of two squares and one cube, the best that was known through the circle method about two squares in this context was that they and four cubes sufficed to represent all large numbers. Not so long ago, however, Linnik [15] proved by his ergodic method that, if $\nu(n)$ is the number of representations of n as the sum of two squares and three non-negative cubes, then

$$\nu(n) > n^{2/3-\varepsilon}$$

for $n > n_0$, thus shewing that all large numbers are expressible in the proposed manner. But this work neither supplied an asymptotic formula for $\nu(n)$ nor even shewed that $\nu(n)$ was of the expected order of magnitude. We therefore propose to sketch briefly how we proved the asymptotic formula

$$\nu(n) \sim \frac{1}{27} \pi \Gamma^3\left(\frac{1}{3}\right) \mathfrak{S}(n) n,$$

where $\mathfrak{S}(n)$ is the singular series, thus demonstrating that the theory for two squares and three cubes conforms to the traditional pattern of results in Waring's problem [8]. Note that this is only the third genuine example of an asymptotic formula in Waring's problem where the cardinality of representations of n does not essentially exceed n in order of magnitude (the other two are the explicit formulae for three squares and for four squares), a fact that is related to our avoidance of the circle method.

The source of the method is that, if $r(\mu)$ denotes the number of ways of expressing μ as the sum of two squares, then

$$r(\mu) = 4\sum_{l|\mu}\chi(l) \quad (\mu \neq 0) \tag{6}$$

and

$$\nu(n) = \sum_{X^3+Y^3+Z^3\leqslant n}' r(n-X^3-Y^3-Z^3). \tag{7}$$

Substituting (6) in (7), we find that $\nu(n)$ is expressible as a combination of sums such as

$$\sum_{\substack{X^3+Y^3+Z^3\leqslant n\\ X^3+Y^3+Z^3\equiv n, \bmod k}} 1,$$

where $k \leqslant n^{1/2}$. The latter sums in turn can be evaluated by complicated transformations in terms of the exponential sums

$$\sum_{\substack{X^3+Y^3+Z^3\equiv n, \bmod k\\ 0<X,Y,Z\leqslant k}} e^{2\pi i(aX+bY+cZ)/k},$$

to which our estimates through Deligne's theory are applicable. The formulae thus obtained almost, but not quite, suffice, a very complicated argument involving a deep theory of elliptic curves over finite fields being needed to complete the proof.

A somewhat surprising lacuna in the theory of these mixed problems has been the absence of known asymptotic formulae for the representations of numbers as the sum of three squares and a non-negative k-th power when k is greater than 2. Notwithstanding the existence of an exact formula for the number of ways of expressing a number as the sum of three squares, this question turns out to be unexpectedly difficult for the larger values of k, and it is only now that the asymptotic formulae have been derived by the author by exploiting relatively recent developments in the theory of the Dirichlet's L-functions [11].

We next consider the classical Diophantine equation

$$X^h+Y^h = Z^h+W^h \quad (h>2), \tag{8}$$

which was studied, in particular, by Fermat and Euler. Although these scholars obtained rational parametric solutions when h is 3 or 4, it has been conjectured that the equation has no non-trivial solutions whenever

$h \geqslant 5$. This speculation being obviously extraordinarily difficult to treat in view of its connection with Fermat's Last Theorem, it is of interest to ponder some associated questions involving the expression of a number as the sum of two h-th powers whose resolution would provide some guidance about the matter. Let $r_h(n)$ be the number of ways of expressing n as the sum of two h-th powers (positive or negative, order being relevant), let $N_h(x)$ be the number of positive integers n not exceeding x for which $r_h(n) > 0$, and let $\nu_h(x)$ be the number of those integers for which $r_h(n) > 2$, noting that $\nu_h(x) = 0$ for $h \geqslant 5$ if the conjecture is true. Then we have been able to shew that

$$N_h(x) \sim A(h)x^{2/h} \tag{9}$$

and

$$\nu_h(x) = O(x^{5/(3h-1)+\varepsilon}),$$

thus demonstrating that it is certainly exceptional for a number expressible in the given form to be thus expressible in more than essentially one way. This goes some way in the required direction and is actually true for all $h \geqslant 3$, although so far we have only supplied the full details for the case where h is odd ([6], [7], in which are supplied references to relevant earlier writings by Erdös, Mahler, Greaves, and the author). This work also furnishes an analytic theory of the representation of numbers as the sum of two h-th powers for $h > 2$, a theory that is seen to contrast markedly with the classical theory for the case $h = 2$.

Considerations relating to the density of representations, to which we have previously alluded, preclude the application of the circle method to the additive equation (8), which indeed is even beyond the theoretical powers of that method whenever $h > 3$. Yet we cannot tarry long enough to describe our method in detail on account of its complication. It must therefore suffice to indicate briefly the ideas involved by referring to the case $h = 3$, in which (8) takes the form

$$r(r^2+3s^2) = \varrho(\varrho^2+3\sigma^2) \tag{10}$$

after a simple transformation. Since only solutions with $X+Y \neq Z+W$ serve to give a bound for $\nu_3(x)$, we are led to study (10) subject to $r < \varrho$ and other appropriate conditions. Now (10) is contained in the equation

$$r(r^2+3s^2) = \varrho l, \tag{11}$$

which, being of the form

$$r(r^2+3s^2) \equiv 0 \bmod \varrho,$$

can be studied with great accuracy by the theory of exponential sums in a manner akin to that used in the two squares and three cubes problem. The cardinality of solutions of (11) is too large, however, and it is necessary to take into account the special nature of the number l by means of a sieve method that exploits the idea that, for any prime p, a square σ^2 is *not* a quadratic non-residue, mod$p$. The calculations involved in this refinement are somewhat complicated and involve our estimates for multiple exponential sums of type (5) for $r = 3$.

There is an application of these ideas to the study of the number $\varrho(n)$ of representations of n as the sum of four non-negative cubes. It being at present impossible to find an asymptotic formula or even a positive lower bound for $\varrho(n)$, it is not without interest to elicit as keen an upper bound as possible for $\varrho(n)$. Here our method gives [5]

$$\varrho(n) = O(n^{11/18+\varepsilon}),$$

which represents an improvement on the trivial bound $O(n^{2/3+\varepsilon})$.

It had been guessed by Davenport and others that there is a positive density of numbers expressible as the sum of two cubes of rational numbers, and this was proved by Stephens [17] on the assumption that the Birch–Swinnerton-Dyer conjectures for certain elliptic curves are true. At the level of unconditional results, if $M(x)$ is the number of positive integers up to x that are the sum of two rational cubes, then our result (9) gives

$$M(x) \geqslant N_3(x) > A_1 x^{2/3}$$

with an explicit value for A_1. But our method can be adapted to take meaningful account of the changed circumstances with the consequence that we can shew that [12]

$$M(x) > A_2 x^{2/3} \log x.$$

The calculations involved also shed other light on the conjecture and suggest that it can only be true if the elliptic equation

$$X^3+Y^3 = nZ^3 \tag{12}$$

frequently has a smallest solution in which Z is almost exponentially large in terms of n.

Another interesting question in this field is whether a polynomial $f(x)$ equalling a sum of two integral h-th powers for every integer x is identically of the form

$$\{f_1(x)\}^h + \{f_2(x)\}^h. \tag{13}$$

Our method cannot so far resolve this matter but can at least shew that such polynomials $f(x)$ have certain properties that are consistent with their having the proposed form (13). We should also observe that Schinzel [16] has actually shewn that the answer is in the affirmative provided that certain far-reaching generalizations of the prime-twins conjecture are true.

Our thesis has tended to shew up certain shortcomings in the powerful circle method in spite of the suggestions we have made concerning its improvement. Apart from theoretical limitations, these deficiencies fall into a number of categories. For example, the method is in some respects not very flexible in adapting itself to the peculiar circumstances of individual problems, a penalty no doubt of the wide ranging scope of the machinery. Moreover, for the deeper problems the analysis becomes very complicated, a situation that is brought about in part by the need to consider exponential sums at arguments other than the arithmetically natural values h/k. In view of these facts and our present inability to make further substantial progress with Waring's problem through the circle method, it seems worthwhile to devise an alternative method of some generality that might incorporate some of the features of the special methods already mentioned. We therefore go on to describe a procedure developed by the author [9] that is applicable in principle to Waring's problem for any exponent and that has already been successful in isolating some new results. In some respects it has a potential for further refinement that is denied the circle method, although we have not succeeded in using it to resolve any of the deeper unsettled questions. Furthermore, the method has nowhere the same universality as that of the circle method.

We hint at the method by considering its relevance to problems involving the equation

$$l^2 + \varphi(l_1, \ldots, l_r) = n, \tag{14}$$

in which there is always a square present and in which $\varphi(l_1, \ldots, l_r)$ is a sum of powers. The underlying idea, implemented in practice with rather more refinement than our remarks here might suggest, is to split up $\varphi(l_1, \ldots, l_r)$ into two sums of powers $\varphi_1(l_1, \ldots, l_s)$, $\varphi_2(l_{s+1}, \ldots, l_r)$ in a suitable way and to prove that the expected asymptotic expression $L(m)$ for the number $N(m)$ of solutions of

$$l^2+\varphi_1(l_1, \ldots, l_s) = m$$

is in fact always valid save possibly for a small exceptional set of m. If this can be achieved, then one can estimate the number of solutions of (14) by considering

$$\sum N\{n-\varphi_2(l_{s+1}, \ldots, l_r)\}$$

provided that $r-s$ is not too small.

The connection between $N(m)$ and $L(m)$ is treated by attempting to shew that the variance

$$\sum_{m\leqslant x}\{N(m)-L(m)\}^2$$

is small, to which end we require a good asymptotic formula for

$$\sum_{m\leqslant x} N^2(m).$$

Now the latter sum is obviously equal to the number of solutions in certain integers of the equation

$$l^2-\lambda^2 = \varphi_1(\lambda_1, \ldots, \lambda_s)-\varphi_1(l_1, \ldots, l_s) = \psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s),$$

say, and hence of

$$\varrho\sigma = \psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s),$$

where in particular ϱ, σ are of the same parity. For given ϱ, this gives rise to the condition

$$\psi(\lambda_1, \ldots, \lambda_s, l_1, \ldots, l_s) \equiv 0 \bmod \varrho,$$

which can be treated by means of the exponential sums

$$\sum_{\lambda_1,\ldots,\lambda_s,l_1,\ldots,l_s} e^{2\pi i\psi(\lambda_1,\ldots,\lambda_s,l_1,\ldots,l_s)h/k} \qquad (k|\varrho)$$

by a variation of earlier methods described. The analysis is then completed by using, inter alia, many of the properties of these sums that were previously developed in connection with the circle method, it being notable that we now need only work with trigonometrical sums corresponding to rational arguments.

We notice that our procedure consists partly of reducing our problem to another one in which one of the unknowns occurs linearly. If the lowest exponent occurring in the given problem is greater than two, then several transformations are needed to secure a linear problem and the details can become very formidable. However, a simple proof of the asymptotic formula in the nine cubes problem can be derived in this manner.

When examined systematically, our method is seen to have many links with the circle method in spite of the different genesis, the occurrence of similar exponential sums being a case in point. But the exponential sums in our method are shorn of arithmetically irrelevant analytic complications, thus lightening the potential task of effecting Kloosterman type refinements when these might be relevant or possible.

Our mention of the sizes of the solutions of the Diophantine equation (12) gives us an opening to introduce our final topic. This is the Pellian equation

$$T^2 - DU^2 = 1,$$

whose fundamental solution $\eta_D = T + \sqrt{D}U$ is known to satisfy

$$2\sqrt{D} < \eta_D < e^{A\sqrt{D}\log D}$$

for positive (non-square) determinants D. Since these inequalities have more or less represented the full extent of our knowledge, the author [13] has evolved a lattice point method that determines the distribution of the determinants D for which η_D is limited by small functions of D. Although the results obtained can only be rigorously substantiated for the smaller limits, the author in fact believes they are true for much larger limits. If we were right in this opinion, and in our reasons for holding it, then some interesting facts concerning the class number $h(D)$ of properly primitive indefinite binary quadratic forms

$$ax^2 + 2bxy + cy^2$$

of determinant $D = b^2 - ac$ would emerge. For example, we could obtain the asymptotic formula

$$\sum_{D \leqslant x} h(D) \sim (25/12\pi^2)x\log^2 x,$$

which would settle a matter that has been open since it was first raised by Gauss in the Disquisitiones Arithmeticae (V, Art. 304). As it is, we can obtain unconditional lower bounds for the above sum that advance our knowledge. Moreover, we are led to conjecture that

$$\sum_{\substack{p\leqslant x\\ p\equiv 1, \bmod 4}} h(p) \sim \frac{x}{6}$$

and that, if $\tau(\beta, x)$ is the number of determinants $p \equiv 1 \bmod 4$ for which $h(p) > \beta$, then

$$\lim_{x\to\infty} \frac{\tau(\beta, x)}{x/\log x} \sim \frac{1}{3\beta} \tag{15}$$

as $\beta \to \infty$. Impressive corroboration of these ideas comes from the present work of Henri Cohen, who simultaneously has been led by entirely different considerations of a more algebraic nature to enunciate conjectures about the behaviour of $h(p)$. His work, which is strongly supported by numerical evidence, agrees with ours in all areas where the subjects of investigation coincide (in particular, equation (15)), although it should be stressed that he and the author by no means study the same questions overall. Conditional work on similar matters has also been described in a recent paper by Takhtajan and Vinogradov [18]. As with the earlier matters discussed, this topic shews there is much life left in many of the important questions in number theory that were first raised centuries ago.

References

[1] Davenport H., Cubic Forms in Thirty Two Variables, *Proc. Trans. Roy. Soc London Ser. A* **251** (1959), pp. 193–232.

[2] Davenport H., Cubic Forms in 29 Variables:, *ibid.* **266** (1962), pp. 287–298.

[3] Davenport H., Cubic Forms in Sixteen Variables, *ibid.* **272** (1963), pp. 258–303.

[4] Heath-Brown D. R., Cubic Forms in 10 Variables, to appear.

[5] Hooley C., On the Representations of a Number as the Sum of Four Cubes, *Proc. London Math. Soc.* **36** (3) (1978), pp. 117–140.

[6] Hooley C., On the Numbers that Are Expressible as the Sum of Two Cubes, *J. Reine Angew. Math.* **314** (1980), pp. 146–173.

[7] Hooley C., On Another Sieve Method and the Numbers that Are a Sum of Two h-th Powers, *Proc. London Math. Soc.* **43** (3) (1981), pp. 73–109.

[8] Hooley C., On Waring's Problem for two Squares and Three Cubes, *J. Reine Angew. Math.* **328** (1981), pp. 161–207.

[9] Hooley C., On a New Approach to Various Problems of Waring's Type, *Recent Progess in Analytic Number Theory,* Vol. 1, Academic Press, 1981.

[10] Hooley C., On Exponential Sums and Certain of Their Applications, Journées Arithmétiques 1980, *London Math. Soc. Lecture Notes Series* **56**, Cambridge, 1982.

[11] Hooley C., *On Waring's Problem for Three Squares and an h-th Power*, to appear.

[12] Hooley C., *On the Numbers that Are a Sum of Two Rational Cubes*, to appear.
[13] Hooley C., *On the Pellian Equation and the Class Number of Indefinite Binary Quadratic Forms*, to appear.
[14] Katz N. M., Sommes exponentielles, *Astérisque* **79**, Société Mathématique de France (1980).
[15] Linnik Ju. V., Additive Problems Involving Squares, Cubes, and Almost Primes, *Acta Arith.* **21** (1972), pp. 413–422.
[16] Schinzel A., On the Relation between Two Conjectures on Polynomials, *Acta Arith.* **38** (1982), pp. 285–322.
[17] Stephens N. M., A Corollary to a Conjecture of Birch and Swinnerton-Dyer, *J. London Math. Soc.* **43** (1968), pp. 146–148.
[18] Takhtajan I. A., and Vinogradov A. I., The Gauss–Hasse Hypothesis on Real Quadratic Fields with Class Number One, *J. Reine Angew. Math.* **335** (1982), pp. 40–86.

DEPARTMENT OF PURE MATHEMATICS
UNIVERSITY COLLEGE
CARDIFF, GREAT BRITAIN

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

WU-CHUNG HSIANG

Geometric Applications of Algebraic K-theory

Algebraic K-theory has been one of the most important mathematical developments of the last two decades. In the reports [81], [68], [6] of past International Congresses of Mathematicians, emphasis was given to the algebraic aspects of the theory. In this report, I shall concentrate on its geometric applications. After all, the theory was initiated by Reidemeister [70], Franz [38], de Rham [24], and J. H. C. Whitehead [89] (see also [90]) who introduced in the 30's some invariants for solving geometric problems. The revival of the interest in these invariants in the early 60's, which were the seeds of algebraic K-theory (Milnor [62], Smale [76] and Kervaire's exposition on the S-cobordism theorem of Barden–Mazur–Stallings [54]) also arose from geometric considerations. At the end of this note, I shall make a few conjectures.

Due to limitation of space, I skip the Hermitian K-theory and Novikov's conjecture on higher signatures of closed aspherical manifolds altogether.

Some of the geometric problems dealt with here have very interesting and equally important Hermitian analogues. The interested readers might consult [30], [34], [88].

I. $K_1(A)$, $Wh_1(\pi)$, simple homotopy type and h-cobordism

Let A be an associative ring with unit 1. The group of all non-singular $n\times n$ matrices over A will be denoted by $\mathrm{GL}(n, A)$. Identifying each $M\in \mathrm{GL}(n, A)$ with $\begin{bmatrix} M & 0\\ 0 & 1\end{bmatrix}\in \mathrm{GL}(n+1, A)$, we obtain inclusions $\mathrm{GL}(1, A)\subset \ldots \subset \mathrm{GL}(n, A)\subset \ldots$ The union is called the infinite *general linear group* $\mathrm{GL}(A)$. A matrix is *elementary* if it coincides with the identity matrix except for one off-diagonal entry. It was observed by J. H. C.

Whitehead [4, p. 226], [63, p. 359] that the subgroup $E(A) \subset \mathrm{GL}(A)$ generated by all elementary matrices is a perfect group and is precisely equal to the commutator subgroup of $\mathrm{GL}(A)$. We define

$$K_1 A = \mathrm{GL}(A)/E(A) \tag{1}$$

which may be viewed as a generalization of the determinant function for matrices. Let π denote a multiplicative group and $Z[\pi]$ the corresponding integral group ring. We have natural inclusions $\pm\pi \subset \mathrm{GL}(1, Z[\pi]) \subset \mathrm{GL}(Z[\pi])$, where $\pm\pi$ denotes the subgroup of (1×1)-matrices $(\pm g)$, $g \in \pi$. The cokernel $K_1(Z[\pi])/\mathrm{image}(\pm\pi)$ is called the *Whitehead group* $Wh_1(\pi)$. Clearly, $K_1(A)$ and $Wh_1(\pi)$ are covariant functors of rings and groups to Abelian groups respectively.

Whitehead [90] introduced the notion of *simple homotopy* which is finer than *homotopy*. Let L_0 and L_1 be finite CW-complexes such that L_1 is obtained from L_0 by attaching a k-cell e_k to L_0 along a $(k-1)$-cell $e^{k-1} \subset \partial e^k$. Call this procedure *simple expansion* and the reverse procedure *simple collapsing*. Simple homotopy is the equivalence relation generated by simple expansion and simple collapsing. Let X and Y be the underlying topological space of the CW-complexes L and K, and let $f\colon X \to Y$ be a homotopy equivalence. Using the CW-complex structures of L and K, we may homotope f to a cellular map $g\colon L \to K$. By introducing the mapping cylinder

$$M_g = X \times [0,1] \cup Y/\{(x,1) = g(x) \mid x \in X\} \tag{2}$$

we obtain a CW-complex pair (M_g, L) such that L is a deformation retract of M_g. It is not difficult to see that the inclusion $K = K \times 0 \subset M_g$ is a simple homotopy equivalence, we shall say that f is simple if $L \subset M_g$ is simple. It was proved in [63, pp. 378–384] and [21] that this definition only depends on the underlying spaces X, Y and the map f, i.e., it is independent of the CW structures L and K of X and Y, and the map g.

Using simple expansions and simple collapsings repeatedly, we may replace (M_g, L) by a CW-complex pair (K_1, L_1), satisfying the following conditions:

(a) (M_g, L) are simply homotopic to (K_1, L_1) respectively and $L_1 \subset K_1$;

(b) K_1 arises from L_1 by attaching a finite number of k-dim cells $\{e_i^k\}$ and $(k+1)$-dim cells $\{e_j^{k+1}\}$ for $k \geqslant 2$. (3)

Consider the universal covering complexes $\hat{L}_1 \subset \hat{K}_1$ of $L_1 \subset K_1$. The fundamental group π will be identified with the group of covering transformations, so that each $\sigma \in \pi$ determines a mapping

$$\sigma\colon (\hat{K}_1, \hat{L}_1) \to (\hat{K}_1, \hat{L}_1), \tag{4}$$

which is *cellular*. If $C_*(\hat{K}_1, \hat{L}_1)$ denotes the cellular chain complex, then each $\sigma \in \pi$ determines a chain map

$$\sigma_\#\colon C_*(\hat{K}_1, \hat{L}_1) \to C_*(\hat{K}_1, \hat{L}_1) \tag{5}$$

and this action makes the chain group $C_p(\hat{K}_1, \hat{L}_1)$ into a free $Z[\pi]$-module with a basis obtained by making a choice of a lift to $\hat{K}_1$ of each p-cell of $K_1 - L_1$. Therefore, we have an isomorphism

$$0 \to C_{k+1}(\hat{K}_1, \hat{L}_1) \xrightarrow{\hat{d}_{k+1}} C_k(\hat{K}_1, \hat{L}_1) \to 0 \tag{6}$$

of free $Z[\pi]$-modules with the liftings $\{\hat{e}_j^{k+1}\}$ and $\{\hat{e}_i^k\}$ of $\{e_j^{k+1}\}$ and $\{e_i^k\}$ as bases. Using these bases, $\hat{d}_{k+1}$ determines an element in $\mathrm{GL}(Z[\pi])$ and thus an element $\tau(K_1, L_1)$ in $Wh_1(\pi)$. It was proved in [63] and [54] that the torsion $\tau(K_1, L_1)$ is independent of the choices[1] and it only depends on f. Denote it by $\tau(f) \in Wh_1(\pi)$. Let us summarize these fatcs in the following theorem.

THEOREM 1.1. *Let X, Y be the underlying topological spaces of the* CW-*complexes K and L, and let $f\colon X \to Y$ be a continuous map. Let $g\colon K \to L$ be a cellular map homotopic to f. Then f determines an element $\tau(f) \in Wh_1(\pi)$ depending only on $f\colon X \to Y$ such that g is simple if and only if $\tau(f) = 0$.*

Applying simple homotopy theory to manifolds, let us consider the following geometric problem. Let $(W^{n+1}; M_0^n, M_1^n)$ be a triad of compact manifolds such that $\partial W^{n+1} = M_0^n \cup M_1^n$. We say that W^{n+1} is an *h-cobordism* between M_0^n and M_1^n if M_i^n $(i = 0, 1)$ are deformation retracts of W^{n+1}. The simplest example of an h-cobordism is $(W^{n+1} = M^n \times [0, 1]$; $M_0^n = M^n \times 0$, $M_1^n = M^n \times 1)$. If $(W^{n+1}; M_0^n, M_1^n)$ is a smooth h-cobordism (i.e., W^{n+1} is a smooth manifold), $\pi_1(W^{n+1}) = 1$ and $n \geqslant 5$, then the remarkable theorem of Smale [76], [64] asserts that W^{n+1} is diffeomorphic to $M_0^n \times [0, 1]$ (and also to $M_1^n \times [0, 1]$). Our interest is focused on the case $\pi = \pi_1 W^{n+1} \neq 1$. If $(W^{n+1}; M_0^n, M_1^n)$ is a smooth h-cobordism,

---

[1] Since we may have different liftings of the cells, we pass from $K_1(Z[\pi])$ to $Wh_1(\pi)$ in order to make the invariant well-defined.

then a C^1-triangulation $t\colon (K; L_0, L_1)\to(W^{n+1}; M_0^n, M_1^n)$ gives rise to a combinatorial cobordism which has a handlebody structure from the triangulation [63]. Or, if $(W^{n+1}; M_0^n, M_1^n)$ is a topological h-cobordism and if $n\geqslant 5$, then (W^{n+1}, M_0^n) has a handlebody decomposition [56]. By a handlebody structure of W^{n+1} on M_0^n, we mean a filtration

$$Y^{(0)} = M_0^n \subset Y^{(1)} \subset \ldots \subset Y^{(l)} = W^{n+1} \tag{7}$$

such that

$$\begin{aligned}&\text{for each } i>0, \text{ there is an embedding}\\ &f_i\colon S^{j_i-1}\times D^{n-j_i+1}\to Y^{(i-1)}, \text{ and a homeo-}\\ &\text{morphism } Y^{(i)}\to Y^{(i-1)}\cup_{f_i} D^{j_i}\times D^{n-j_i+1} \text{ rel } Y^{i-1}.\end{aligned} \tag{8}$$

The union of a k-handle and a $(k+1)$-handle $J = D^k\times D^{n-k+1}\cup D^{k+1}\times\times D^{n-k}$ along $D^k\times H_1 = H_2\times D^{n-k}$, where $H_1\subset \partial D^{n-k+1}$ and $H_2\subset\partial D^{k+1}$ are codim 0 discs, will be called a *trivial pair of handles*. Introducing and cancelling a trivial pair of handles correspond to elementary expansion and elementary collapsing in simple homotopy theory. Following these procedures, if $n\geqslant 5$, we may produce a handlebody structure of (7) such that $j_1 = \ldots = j_k = k$ and $j_{k+1} = \ldots = j_{l-1} = k+1$ $(2\leqslant k\leqslant n-3)$; i.e., W^{n+1} is obtained from M_0^n by attaching k-handles and $(k+1)$-handles. By making a choice of lifting of the cores[2] of the handles to $(\hat{W}^{n+1}, \hat{M}_0^n)$, the universal covers of the pair (W^{n+1}, M_0^n), we get a chain complex

$$0\to C_{k+1}(\hat{W}^{n+1}, \hat{M}_0^n)\xrightarrow{\hat{d}_{k+1}} C_k(\hat{W}^{n+1}, \hat{M}_0^n)\to 0 \tag{9}$$

of based $Z[\pi]$-modules as in (6). The map $\hat{d}_{k+1}$ determines an element $\tau\in Wh_1(\pi(W^{n+1}))$ which is the torsion[3] of the inclusion of $M_0^n\subset W^{n+1}$. This element only depends on the underlying topological space and not on the handle structure of (W^{n+1}, M_0^n) and it is denoted by $\tau(W^{n+1}, M_0^n)$.

In correspondence to Theorem 1.1, we have the following S-cobordism theorem due to D. Barden, B. Mazur and J. Stallings [3], [54], [63].

---

[2] I.e., $f(S^{j_i-1}\times 0)$ of (8) where 0 is the center of D^{n-j_i+1}.

[3] Note the asymmetry of M_0, M_1 in the definition. If we wish to consider (W, M_1), we have to consider the duality of [63, pp. 393–398].

THEOREM 1.2 (A) *Let* $(W^{n+1}; M_0^n, M_1^n)$ *be a smooth*, PL *or topological* h*-cobordism for* $n \geq 5$. *Then,* W^{n+1} *is diffeomorphic*, PL *homeomorphic or homeomorphic to* $M_0^n \times [0,1]$ *if and only if* $\tau(W^{n+1}, M_0^n) = 0$.

(B) *For a given manifold* M_0^n $(n \geq 5)$ *and an element* $\tau_0 \in Wh_1(\pi_1(M_0^n))$, *there exists an* h*-cobordism* $(W^{n+1}; M_0^n, M_1^n)$ *such that* $\tau(W^{n+1}, M_0^n) = \tau_0$ *and* W^{n+1} *is smooth or* PL *if* M_0^n *is so.*

II. Higher K-groups and pseudo-isotopy theory

For A a ring with unit 1, we observed in § I that $E(A) = [\mathrm{GL}(A), \mathrm{GL}(A)] \subset \mathrm{GL}(A)$ is perfect. Let $\mathrm{BGL}(A)$ be the classifying space of $\mathrm{GL}(A)$. Construct $\mathrm{BGL}^+(A)$, Quillen's "+" construction [67], from $\mathrm{BGL}(A)$ by attaching 2-cells and 3-cells such that there is a homology equivalence $\mathrm{BGL}(A) \to \mathrm{BGL}^+(A)$ over Z, and $\pi_1 \mathrm{BGL}^+(A) = \mathrm{GL}(A)/E(A) = K_1(A)$. In fact, $\mathrm{BGL}^+(A)$ is an infinite loop space. $K_i(A)$ *is defined to be* $\pi_i \mathrm{BGL}^+(A)$ $(i \geq 1)$. Waldhausen [84] generalized this definition to construct higher algebraic K-groups of a pointed connected CW-complex X as follows. Let G be the loop group[4] of X and let $R = \Omega^\infty S^\infty[G_+]$ be the "group ring"[5] of G over $\Omega^\infty S^\infty$. Form the matrix ring $M_n(R)$. Consider the pull-back diagram

$$\begin{array}{ccc} \hat{\mathrm{GL}}_n(R) & \longrightarrow & M_n(R) \\ \downarrow & & \downarrow{\scriptstyle \pi_0} \\ \mathrm{GL}_n(\pi_0 R) & \longrightarrow & M_n(\pi_0 R) \end{array} \tag{10}$$

and let $\hat{\mathrm{GL}}(R) = \varinjlim_{n} \hat{\mathrm{GL}}_n(R)$. It turns out that $\mathrm{B}\hat{\mathrm{GL}}(R)$ exists, and using the fact that $[\pi_0 \hat{\mathrm{GL}}(R), \pi_0 \hat{\mathrm{GL}}(R)] = E(Z[\pi_0 G])$ we may perform the "+" construction for $\mathrm{B}\hat{\mathrm{GL}}(R)$ such that $\mathrm{B}\hat{\mathrm{GL}}^+(R)$ is an infinite loop space [84], [79]. Waldhausen defined

$$A(X) = \mathrm{B}\hat{\mathrm{GL}}^+(R) \tag{11}$$

and $K_i(X) = \pi_i A(X)$ for $i \geq 1$.

---

[4] G is either a simplicial group model or a topological group model of the loop space ΩX of X.

[5] Since $\Omega^\infty S^\infty$ is not a topological ring, R is only a group ring in an appropriate sense. This causes most of the technical difficulties.

Based on this model, some computations for $\pi_i(A(X))\otimes Q$ were given in [49], [35], [15]. Invariant theory plays an important rôle in these results.

Due to the fact that $\Omega^\infty S^\infty$ is not an honest (topological) ring, Waldhausen's original way of introducing $A(X)$ is different from the above. See [84], [78], [79] for the details.

$A(X)$ is closely related to pseudo-isotopy theory. Let us recall some developments before the publication of [84]. If M is a compact differentiable manifold (generally with boundary), a pseudo-isotopy of M is a diffeomorphism $f\colon M\times I\to M\times I$ such that $f|M\times 0 = \mathrm{id}$. Let $P(M)$ denote the space of pseudo-isotopies endowed with the C^∞-topology. We are interested in computing $\pi_i(P(M))$. This problem was first studied by Cerf [20] and later by Hatcher and Wagoner [42]. Their idea roughly goes as follows. Connect a given pseudo-isotopy $f\colon M\times I\to M\times I$ to the identity pseudo-isotopy $f_0\colon M\times I\to M\times I$ by means of a generic map $F\colon M\times I\times I\to I\times I$ such that $F|M\times I\times 0 = pf_0$ and $F|M\times I\times 1 = pf$ where p denotes the projection to the second I factor. Choosing F carefully, the construction produces a one-parameter family of handlebodies as follows: Let t denote the coordinate of the second I factor. For $t=0$ (resp. $t=1$), $M\times I$ is the given product structure induced by a gradient-like vector field associated to the Morse function pf_0 (resp. pf). There is a finite number of birth points[6] for $0<t<\varepsilon_0$ (ε_0 a small positive number) such that at $t=\varepsilon_0$ the Morse function F_{ε_0} gives rise to "trivial pairs of handles". Similarly, there is a finite number of death points for $1-\varepsilon_1 < t < 1$ (ε_1 a small positive number) such that the handles are cancelled in "trivial pairs" at $t = 1-\varepsilon_1$. The one-parameter family of Morse functions F_t ($\varepsilon_0\leqslant t\leqslant 1-\varepsilon_1$) gives rise to a one-parameter family of handlebodies over the subinterval $[\varepsilon_0, 1-\varepsilon_1]\subset[0,1]$. Based on analysis of the parametrized handlebodies, Hatcher and Wagoner [42] relate $\pi_0 P(M)$ to $Wh_2(\pi_1 M)$, a quotient of $K_2(Z[\pi_1 M])$ for $\dim M\geqslant 6$.

In [40], Hatcher studied the space of PL pseudo-isotopy spaces $P_{\mathrm{PL}}(M)$ and the following stability question:

> Let $P_{\mathrm{PL}}(M)\subset P_{\mathrm{PL}}(M\times I)$ (resp. $P(M)\subset P(M\times I)$) be the natural inclusion essentially given by $f\times\mathrm{id}$. Is $\pi_i(P_{\mathrm{PL}}(M))\to\pi_i(P(M\times I))$ an isomorphism for $i\leqslant\dim M$?

[6] See [20], [42] for the precise definitions.

He claimed the stability theorem for $P_{\mathrm{PL}}(M)$ and then Burghelea and Lashof extended it to $P(M)$ [16]. Unfortunately, there are some flaws in the proof of [40]. Based on his work on pseudo-isotopy by eliminating the higher order singularities [51] and a modification of Hatcher's argument, K. Igusa in a still unpublished paper has proved that the stability theorem is valid for $P(M)$ with a somewhat smaller range. Elaborating on the multi-disjunction lemma, in his thesis, T. Goodwillie has claimed that the stability ranges for $P(M)$ and $P_{\mathrm{PL}}(M)$ are the same (yet unpublished). The interest in the stability theorem stems from the observation that

$$\begin{gathered} \boldsymbol{P}(M) = \lim_i P(M \times I^i), \\ (\text{resp. } \boldsymbol{P}_{\mathrm{PL}}(M) = \lim_i P_{\mathrm{PL}}(M \times I^i)) \end{gathered} \tag{13}$$

becomes an infinite loop space and we can apply homotopy theory and the categorical machinery. This is the starting point for Waldhausen.

Let us study $Wh^{\mathrm{Diff}}(M)$, the double delooping of $\boldsymbol{P}(M)$. Motivated by consideration of the parametrized handlebodies for studying $\pi_0 P(M)$, consider a "rigid handlebody theory", a manifold model of Waldhausen's expansion space [84]. (We follow the exposition of [46].)

Let ∂_0 be an $(n+k-1)$-dim manifold and π_0: $\partial_0 \to \Delta^k$ a differentiable bundle map such that the fibers are $(n-1)$-dim manifolds (generally with boundary). Suppose that ∂_0 is a codim 0 submanifold of ∂Y for a manifold Y and π: $Y \to \Delta^k$ is a bundle projection extending π_0. We say that Y is a *k-parameter family of rigid handlebodies on* ∂_0 if there is a filtration:

$$Y^{(0)} = \partial_0 \subset Y^{(1)} \subset \ldots \subset Y^{(l)} = Y \tag{14}$$

satisfying the following conditions:

(a) For each $i > 0$, there is an embedding

$$f_i\colon S^{j_i-1} \times D^{n-j_i} \times \Delta \to Y^{(i-1)}$$

and a homeomorphism

$$Y^{(i)} \xrightarrow[\simeq]{h_i} Y^{(i-1)} \cup_{f_i} D^{j_i} \times D^{n-j_i} \times \Delta^k$$

rel $Y^{(i-1)}$ such that f_i and d_i preserve the projection onto Δ^k.

(b) $M^{(i)} = Y^{(i)} \cup_{\partial_0 = \partial_0 \times \{1\}} \partial_0 \times I$ is a manifold, even though $Y^{(i)}$ need not be.

(c) Let $\partial_+ Y^{(i)} = \mathrm{cl}\left(\partial M^{(i)} - (\partial_0 \times \{0\} \cup \partial M^{(i)} | \partial \Delta^k)\right)$ where $\partial M^{(i)} | \partial \Delta^k$ is the part lying over $\partial \Delta^k$. Then, $f_i(S^{j_i - 1} \times D^{n-j_i} \times \Delta^k) \subset \partial_+ Y^{(i-1)}$ and f_i is a differentiable embedding into $\partial_+ Y^{(i-1)}$ (after we smooth the corners). Note that h_i has an obvious extension to

$$M^{(i)} = M^{(i-1)} \cup D^{j_i} \times D^{n-j_i} \times \Delta^k,$$

and also assume that this is a diffeomorphism (again after smoothing the corners).

The attaching data, the f_i's and d_i's, are a part of the structure, but independent handles may be attached in any order. We may construct a *category E^n_k which has the k-parameter families of rigid handlebodies as objects and the compositions of isomorphisms and cancelling of trivial pairs of handles*[7] *as morphisms.*

The obvious definitions of face and degeneracy make $E^n_{\bullet}$ a simplicial category and appropriate inclusions and quotients make it into a cofibration category in the sense of [84], [46], [59].

Let X be a space and let ξ be a stable vector bundle over X. The categories $E_k(X, \xi)^n$, $k, n = 0, 1, \ldots$, are defined as follows. The objects are diagrams

$$f\colon (Y, \partial_0) \to X \tag{16}$$

where (Y, ∂_0) is an object of E^n_k and $f\colon Y \to X$ is a continuous map, together with a stable bundle isomorphism $\psi\colon tY \to f^* \xi$ where tY is the stable tangent bundle of Y. The morphisms and cofibrations of E^n_k, appropriately modified with the data on the induced bundles from ξ, define morphisms and cofibrations of $E_k(X, \xi)^n$. Note that $E_k(X; \xi)^n$ has a composition law "$+$" — disjoint union of (Y, ∂_0)'s — and hence the classifying space has an infinite delooping in the sense of Γ-spaces [72]. On the other hand, we can make use of the cofibration structure such that the $S_{\bullet}$ construction of [84] and the Q construction of [66], [46] apply to this situation as explicit deloopings of $E_k(X; \xi)^n$.

Let $E^h_k(X; \xi)^n \subset E_k(X; \xi)^n$ be the full subcategory of objects such that $\partial_0 \subset Y$ is a homotopy equivalence. We can also deloop $E^h_{\bullet}(X; \xi)^n$ by means of $S_{\bullet}$ construction.

[7] For technical reasons, we don't really cancel the trivial pairs geometrically. See [46] for details.

Multiplying by D^1, we define functors

$$
\begin{aligned}
&S_{.}E_{.}(X;\,\xi)^n \xrightarrow{\Sigma} S_{.}E_{.}(X;\,\xi)^{n+1},\\
&S_{.}E^h_{.}(X;\,\xi)^n \xrightarrow{\Sigma} S_{.}E^h_{.}(X;\,\xi)^{n+1}.
\end{aligned} \tag{17}
$$

Set

$$
\begin{aligned}
BE(X;\,\xi) &= \lim_n |S_{.}E_{.}(X;\,\xi)^n|,\\
BE^h(X;\,\xi) &= \lim_n |S_{.}E^h_{.}(X;\,\xi)^n|,
\end{aligned} \tag{18}
$$

where the limit is taken with respect to Σ. It turns out that $BE(X;\,\xi)$ and $BE^h(X;\,\xi)$ are weakly homotopically equivalent to $\Omega^{\mathrm{fr}}(X)$ — the infinite loop space associated to the frame bordism and $Wh^{\mathrm{Comb}}(X)$ of [84], respectively, if X is a finite complex. (They are independent of ξ!) It has been shown that $Wh^{\mathrm{Comb}}(X)$ is rationally equivalent to $Wh^{\mathrm{Diff}}(M)$ if X is homotopy equivalent to M with the tangent bundle of M, $t(M)$ stably equivalent to ξ. (This is the reason why we keep ξ in the construction.) In fact, Waldhausen has recently claimed that $Wh^{\mathrm{Comb}}(X)$ is $Wh^{\mathrm{Diff}}(M)$.

Finally, let us state the remarkable result of Waldhausen [84]. It comes out of his "localization theorem" [84], [59], [83].

THEOREM 2.1 *There is a fibration up to homotopy*

$$\Omega BE(X;\,\xi)\to A(X;\,\xi)\to BE^h(X;\,\xi)$$

which is weakly homotopically equivalent to

$$\Omega^{\mathrm{fr}}(X)\to A(X)\to Wh^{\mathrm{Comb}}(X)$$

if X if a finite complex.

Since $Wh^{\mathrm{Comb}}(M)$ is (at least rationally) equivalent to $Wh^{\mathrm{Diff}}(M)$, one can easily see the importance of the functor $A(X)$ if one is interested in computing $\pi_i\big(P(M)\big) = \pi_{i+2}\big(Wh^{\mathrm{Diff}}(M)\big)$.

III. $K_0(A)$, obstructions to being a finite CW-complex and to finding a boundary for an open manifold

Again let A be a ring with unit 1. Let K_0A be the additive group having one generator, $[P]$, for each finitely generated projective module P over A, and one relation, $[P]-[P_0]-[P_1]$, for a short exact sequence $0\to P_0\to P\to P_1\to 0$. In other words, K_0A is the "Grothendieck group" associated to the category of finitely generated projectives over A. The class of free A-modules of rank 1 generates a cyclic subgroup of $K_0 A$. The quotient

$$K_0 A/(\text{subgroup generated by free modules})$$

is called the *projective class group*, $\tilde{K}_0(A)$. If $A = Z[\pi]$, then $\tilde{K}_0(Z[\pi])$ is sometimes written as $\tilde{K}_0(\pi)$.

Let X be a connected CW-complex and let Y be a connected *finite* CW-complex. We say that X is dominated by Y if there exist $f\colon X \to Y$, $g\colon Y \to X$ such that $fg\colon Y \to Y$ is homotopic to the identity. We would like to know whether X itself is of the homotopy type of a finite complex. It turns out that we may choose Y such that

$$H_i(\hat{M}_i, \hat{X}) = 0$$

for $i \neq k$ ($k \geqslant 2$) and $H_k(\hat{M}_f, \hat{X})$ is a finitely generated projective module over $Z[\pi_1 X]$ where M_f denotes the mapping cylinder of f and $(\hat{M}_f, \hat{X})$ is the universal covering of the pair (M_f, X). The class

$$\sigma(x) = (-1)^k [H_k(\hat{M}_f, \hat{X})] \in \tilde{K}_0(\pi_1 M) \tag{19}$$

turns out to be well-defined, independent of the choice of Y or of the integer k. In [86], the following fundamental theorem was proved:

THEOREM 3.1 (A) *For X a connected CW-complex dominated by a finite complex, X is of the homotopy type of a finite complex iff $\sigma(X) = 0$.*

(B) *Let $\sigma_0 \in \tilde{K}_0(\pi)$ be a given element with π a finitely presented group. There exists a connected CW-complex X dominated by a finite CW-complex with $\pi_1 X = \pi$ and $\sigma(x) = \sigma_0$.*

Let W^n ($n > 5$) be a smooth (or PL) open manifold. If there exists an arbitrarily large compact set with 1-connected complement and if $H_*(W)$ is finitely generated as an Abelian group, then, as was proved in [10], W is the interior of some smooth (or PL) compact manifold, $\overline{W}$. For the general case, L. Siebenmann developed the following theory [74]. Let W^n ($n > 5$) be a connected smooth open manifold and let ε be an end of W^n. We say that ε is *tame* if it satisfies the following conditions:

(A) There exists a sequence of neighborhoods of ε, $U_1 \supset U_2 \supset \dots \supset U_i \supset \dots$ such that $\bigcap U_i = \emptyset$ and $\pi_1(U_i) \leftarrow \pi_1(U_{i+1})$, $i = 1, \dots$ are isomorphisms. We set $\pi = \pi_1(\varepsilon) = \varprojlim \pi_1(U_i)$ and call it π_1 of the end ε. (20)

(B) Each U_i is dominated by a finite CW-complex.

We may ask whether we can add a boundary to W^n at the tame end ε and reduce our problem to the case where W^n has only one end. In fact, we may choose each U_i of (19) to be a manifold with compact boundary ∂U_i such that $\pi_q(\partial U_i) \simeq \pi_1(U_i) \simeq \pi$ and

$$H_j(\hat{U}_i, \partial\hat{U}_i) = 0$$

for $i \neq k$ $(3 \leqslant k \leqslant n-3)$ where $(\hat{U}_i, \partial\hat{U}_i)$ is the universal covering of the pair $(U_i, \partial U_i)$ and $H_k(\hat{U}_i, \partial\hat{U}_i)$ is a finitely generated projective module over $Z[\pi]$. Then, $(-1)^k[H_k(\hat{U}_i, \partial\hat{U}_i)] = \sigma(U_i) \in \tilde{K}_0(\pi)$ is the obstruction defined in Theorem 3.1. Here is Siebenmann's theorem [74].

THEOREM 3.2. (A) $(-1)^k[H_k(\hat{U}_i,\partial\hat{U}_i)] \in \tilde{K}_0(\pi_1\varepsilon)$ *only depends on* ε, *and we denote it by* $\sigma(\varepsilon)$.

(B) *A boundary can be added to* W^n *at* ε *iff* $\sigma(\varepsilon) = 0$.

IV. Künneth formula for algebraic K-theory and its geometric application

Let T be an infinite cyclic group with a generator t, and let $A[T]$ be the *finite Laurent series ring* of A on t, which is just the group ring of T over A. If α is an automorphism of A, we also have the α-twisted finite Laurent series ring $A_\alpha[T]$. (See [32] for details.) Let $\underline{\underline{\mathrm{Nil}}}(A, \alpha)$ be the full subcategory of the category $\underline{\underline{\mathrm{P}}}(A)$ with objects (P, f) where P is a finitely generated projective module over A with f an α *semilinear nilpotent endomorphism.* Let $\mathrm{Nil}_0(A, \alpha)$ be the Grothendieck group of $\underline{\underline{\mathrm{Nil}}}(A, \alpha)$. The "forgetful functor" defined by "forgetting" the endomorphism f defines a homomorphism j: $\mathrm{Nil}_0(A, \alpha) \to K_0(A)$ and we let $\widetilde{\mathrm{Nil}}_0(A, \alpha)$ denote $\mathrm{Ker}\, j$. It is easy to see that we have a natural decomposition

$$\mathrm{Nil}_0(A, \alpha) = \widetilde{\mathrm{Nil}}_0(A, \alpha) \oplus K_0(A). \tag{21}$$

Let I denote the subgroup of K_1A generated by $x - \alpha_* x$ and let $(K_0A)^{\alpha*}$ denote the subgroup of $x \in K_0(A)$ invariant under α_* (induced by α). Thinking of K_1 and K_0 as homology functors of rings, one might guess from the Künneth formula for the homology groups of a space E fibering over S^1, $F \to E \to S^1$, that there should be an exact sequence

$$0 \to K_1(A)/I \to X \to K_0(A)^{\alpha*} \to 0$$

such that $X \simeq K_1(A_\alpha[T])$. This is not true in general, unless A is (right or left) regular [7]. In fact, it was proved in [4, p. 628] that there is a canonical decomposition

$$K_1(A[T]) = K_1(A) \oplus \widetilde{\mathrm{Nil}}_0(A) \oplus \widetilde{\mathrm{Nil}}_0(A) \oplus K_0(A). \tag{22}$$

This was generalized in [32] to give:

$$K_1(A_\alpha[T]) = X \oplus \widetilde{\mathrm{Nil}}_0(A, \alpha) \oplus \widetilde{\mathrm{Nil}}_0(A, \alpha^{-1}), \tag{23}$$

where X fits into an exact sequence $0 \to K_1(A)/I \to X \to K_0(A)^{\alpha *} \to 0$, and we also have a natural projection:

$$p: K_1(A_\alpha[T]) \to K_0(A)^{\alpha *} \oplus \widetilde{\mathrm{Nil}}_0(A, \alpha). \tag{24}$$

For the group $\pi = G \times_\alpha T$, a semi-direct product, consider $A_\alpha[T] = Z[G \times_\alpha T]$ with $A = Z[G]$. By passing to $Wh_1(G \times_\alpha T)$, we have the formula

$$Wh_1(G \times_\alpha T) = \bar{X} \oplus \widetilde{\mathrm{Nil}}_0(A, \alpha) \oplus \widetilde{\mathrm{Nil}}_0(A, \alpha^{-1}), \tag{25}$$

where $0 \to Wh_1(G)/I \to \bar{X} \to \tilde{K}_0(A)^{\alpha *} \to 0$ $(I = \{y = x - \alpha_* x \mid x \in Wh_1(G)\})$. We also have a natural projection

$$\bar{p}: Wh_1(G \times_\alpha T) \to \tilde{K}_0(G)^{\alpha *} \oplus \widetilde{\mathrm{Nil}}_0(Z[G], \alpha). \tag{26}$$

Now consider the following geometric problem. Let M^n $(n \geqslant 6)$ be a closed smooth (PL or topological) manifold. Is M^n a fibration over S^1 with connected F^{n-1} as fiber? If so, we should have a projection $q: M^n \to S^1$ such that the fiber is homotopic to a connected finite complex X with $\pi = \pi_1 M^n = G \times_\alpha T$ being a semi-direct product of $G = \pi_1 X$ by $T = \pi_1 S^1$. For $\pi = Z$ (i.e., $G = 1$), it was proved in [12] that this condition is sufficient. For the general case, we need to find a codim 1 submanifold $F^{n-1} \subset M^n$ representing the homotopy fiber of q satisfying the following conditions:

(A) When we cut M^n open along F^{n-1}, we have an h-cobordism $(W^n; F_0^n, F_1^n)$ with F_0^{n-1}, F_1^{n-1} diffeomorphic to (PL or homeomorphic to) F^{n-1}; (27)

(B) $\tau(W^n, F_0) \in Wh_1(G)$ vanishes.

If only condition (A) is satisfied, we call it an *almost fibration*. It was proved in [28] that the obstruction to being an almost fibration is an element in $\tilde{K}_0(G)^{\alpha *} \oplus \mathrm{Nil}_0(Z[G], \alpha)$.

Generalizing the problem when a space fibers over S^1, let M_1^n be a closed smooth (PL or topological) manifold of $\dim n \geqslant 6$ with $\pi = \pi_1 M = G \times_\alpha T$ and let $F_1^{n-1} \subset M_1^n$ be a connected codim 1 submanifold with $G = \pi_1 F^{n-1}$ corresponding to the subgroup $G \subset \pi$. Let

$$f\colon M_2^n \to M_1^n \tag{28}$$

be a homotopy equivalence. We ask what is the obstruction $O(f)$ to finding an $(n-1)$-dim submanifold $F_2^{n-1} \subset M_2^n$ and a map

$$g\colon (M_2^n, F_2^{n-1}) \to (M_1^n, F_1^{n-1}) \tag{29}$$

such that

$$\begin{aligned} &\text{(A) } g \text{ is a homotopy equivalence of pairs,}\\ &\text{(B) } g^{-1}(F_1^{n-1}) = F_2^{n-1},\\ &\text{(C) the induced map } g\colon M_2^n \to M_1^n \text{ is homotopic to } f. \end{aligned} \tag{30}$$

$O(f)$ is called the *obstruction to splitting f with respect to F_1^{n-1}*.

THEOREM 4.1 [36]. *Assume that (M_1^n, F_1^{n-1}), $f\colon M_2^n \to M_1^n$ $(n \geqslant 6)$ are given as above. Then the obstruction $O(f)$ to splitting f with respect to F_1^{n-1} is equal to $\bar{p}\tau(f)$ where $\tau(f) \in Wh_1(\pi)$ is the torsion of f.*

One should read S. E. Cappell, A Splitting Theorem for Manifolds, *Invent. Math.* **33** (1976), pp. 66–170 for the generalization of the above theorem.

V. Negative K-groups $K_{-i}(A)$ and some of their geometric applications

In [4, pp. 657–674], H. Bass introduced the functor $K_{-i}(A)$ $(i \geqslant 0)$. He observed that the decomposition (22) is functorial and can be written as

$$0 \to K_1(A) \to K_1(A[t]) \oplus K_1(A[t^{-1}]) \to K_1(A[T]) \to K_0(A) \to 0. \tag{31}$$

It is natural to define K_{-i} $(i > 0)$ recursively using the formula

$$K_{-i}(A) = \operatorname{Coker}\{K_{-i+1}(A[t]) \oplus K_{-i+1}(A[t^{-1}]) \to K_{-i+1}A[T]\}. \tag{32}$$

Bass showed that (31) continued to hold with K_0, K_1 replaced by K_{-i} and K_{-i+1} $(i > 0)$ (and he also defined negative Nil groups). For T^n

$=T\times \ldots \times T$, we have the following decomposition formula [4], [1]:

$$\begin{aligned} Wh_1(\pi\times T^n) &= Wh_1(\pi\times T_1\times \ldots \times\hat{T}_i\times \ldots \times T_n)\oplus \\ &\quad \oplus\tilde{K}_0(\pi\times T_1\times \ldots \times\hat{T}_i\times \ldots \times T_n) \text{ (mod nil terms)} \\ &= Wh_1(\pi)\oplus n\tilde{K}_0(\pi)\oplus \\ &\quad \oplus\sum_{i=2}^{n-1}\binom{n}{i}K_{1-i}(Z[\pi])\oplus \\ &\quad \oplus K_{1-n}(Z[\pi]) \text{ (mod nil terms)}. \end{aligned} \tag{33}$$

It was proved in [8], [19] that if π is a finite group, then

$$K_{-i}(Z[\pi]) = 0 \quad \text{for } i\geqslant 2. \tag{34}$$

Let us now turn to a geometric problem. A Top *stratification* of a space X is an increasing family of closed subsets of X, $\{X^n|n\geqslant 1\}$ such that $X^{(-1)} = \text{fi}$, there is a positive integer N such that $X^{(N)} = X$, and for every n, each component of $X^{(n)}-X^{(n-1)}$ is a topological (Top) n-dim manifold without boundary (possibly empty). The stratification is *locally cone-like* if for every $x\in X^{(n)}-X^{(n-1)}$, there exists a compact Top stratified space L and a stratum-preserving open embedding $h\colon R^n\times cL\to X$ such that $h(0, v) = x$ where cL denotes the open cone over L and v is its vertex. The space L is called a *link of* x and h is called a *local chart*. We shall call a space X a Top CS space [2], [75], if it has a locally cone-line stratification. In [2], we define (combinatorial) PL structures on a Top CS space X compatible with the stratification (see [2] for the precise definition), and then study the existence and uniqueness of such structures on a given Top CS space. Of course, these questions are the refined forms of the problem of triangulating a topological space and the Hauptvermutung. The simplest example of a Top CS space is the suspension $\sum M$ of a closed manifold M. In [62], [77] counterexamples to Hauptvermutung were given with $X = \sum M^n$ $(n\geqslant 5)$ as the underlying topological space and with elements in $Wh_1(\pi_1 M)$ as the invariants to distinguish them. [1], [2] globalize the examples of Milnor and Stallings into an obstruction theory such that the obstruction invariants are generally in the subquotients of K_{-i} of the group ring of the links of various strata.

The obstruction theory of [1], [2] has been explicitly applied to the following problem. Let $R_1, R_2\in O(n)$, the group of orthogonal transformations of R^n. We say that R_1, R_2 are *topologically* (resp. *linearly*) *equiv-*

alent if there is a homeomorphism (resp. linear automorphism) $f\colon R^n \to R^n$ such that $f^{-1}R_1f = R_2\colon R^n \to R^n$. The conjecture that the notions of topological and linear equivalence of rotations should be equivalent was stated by de Rham in 1935 [24] and he reduced it to the case where the rotations have finite order. Note that f induces a homeomorphism

$$h\colon X_1 \to X_2, \tag{34}$$

where $X_i = R^n/\langle R_i\rangle$, the quotient space of R^n by the finite subgroup of $O(n)$ generated by R_i, $i = 1, 2$. Given X_i, the preferred PL structures induced from the rotation R_i, we may then try to deform h to a PL homeomorphism. This is the problem studied in [2]. (See also [69].) Modifying the topologically equivalent R_1, R_2 to new ones, R_1', R_2', if necessary, we manage to kill most of the obstructions in the subquotients of K_{-i} and then apply a version of G-signature theorèm to obtain the following result [47].

Let $R_1, R_2 \in O(n)$ have order $k = l2^m$ where l is odd and $m \geqslant 2$. Suppose that (a) *R_1 and R_2 are topologically equivalent, and* (b) *the eigenvalues of R_1^l and R_2^l are either* 1 *or primitive 2^m-th roots of unity. Then R_1 and R_2 are linearly equivalent.*

If k is odd, then condition (b) is superfluous. In this case, it was proved independently, by Madsen and Rothenberg [60] using a different method from [47]. However, the K_{-i} groups of [1], [2] (see also [57]) still play an important rôle in their work.

The interest of de Rham's problem was revived [55] and there are remarkable counterexamples of this conjecture in [18] if $k = l2^m$, $m \geqslant 2$, $l \neq 1$, and the above condition (b) is not satisfied.

VI. Concluding remarks and some conjectures

One of the problems in algebraic K-theory is to compute $K_i(A)$ ($-\infty < i < \infty$). Emphasizing the geometric applications, we are mostly interested in the case of $A = Z[G]$ for G a finitely presented group. Most algebraic calculations have been carried out for G finite. Let me pose some conjectures about the case when G is *not necessarily* finite or *torsion-free*. In fact, I believe that these problems are more geometrically interesting and they should serve as guide posts for future development.

CONJECTURE 1. *Let G be a finitely presented group. Then $K_{-i}(Z[G]) = 0$ for $i \geqslant 2$. At least, $K_{-i}(Z[G]) = 0$ for $i \gg 0$.*

Before I state the next conjecture, let me single out a class of infinite groups. We say that a closed manifold M^n is a $K(\Gamma,1)$-manifold (an aspherical manifold) if $\pi_i(M^n)=0$ for $i>1$ and $\pi_1 M^n=\Gamma$. Note that Γ is *necessarily torsion-free.*

CONJECTURE 2. *Let Γ be the fundamental group of a closed $K(\Gamma,1)$-manifold. Then* $Wh_1(\Gamma)=\tilde{K}_0(\Gamma)=K_{-i}(Z[\Gamma])=0$ $(i\geqslant 1)$. (See [31] for supporting evidence.)

It is clear that the following conjecture is much stronger than Conjecture 2.

CONJECTURE 3. *Let Γ be a torsion-free group such that $B\Gamma$ has the homotopy type of a finite* CW-*complex. Then* $Wh_1(\Gamma)=\tilde{K}_0(\Gamma)=K_{-i}(Z[\Gamma])=0$ $(i\geqslant 1)$.

For the higher K-groups, let us consider the map of [58]:

$$\lambda_*\colon h_*\big(BG;\underline{\underline{K}}(Z)\big)\to K_*(Z[G]),\tag{35}$$

where $h_*\big(BG;\underline{\underline{K}}(Z)\big)$ denotes a generalized homology theory with coefficients in the spectrum of the algebraic K-theory of Z.

CONJECTURE 4. *If Γ is a torsion-free group such that $B\Gamma$ is of the homotopy type of a finite* CW-*complex, then*

$$\lambda_*\otimes \mathrm{id}\colon h_*\big(B\Gamma;\underline{\underline{K}}(Z)\big)\otimes Q\to K_*(Z[\pi])\otimes Q$$

is an isomorphism.

For $B\Gamma$ having the homotopy type of an aspherical manifold, Conjecture 4 was verified in some special cases [31]. As we pointed out in [35], Conjecture 2 is the algebraic K-theory analogue of Novikov's conjecture on higher signatures. (So are Conjectures 2, 3!) Interested readers should consult [30], [34], [88] for further details about this conjecture.

References

(A suggested list)

[1] Anderson D. R. and Hsiang W.-C., The Functor K_{-i} and Pseudo-Isotopies of Polyhedra, *Ann. of Math.* **105** (1977), pp. 201–233.

[2] Anderson D. R. and Hsiang W.-C., Extending Combinatorial PL Structures on Stratified Spaces, *Invent. Math.* **32** (1976), pp. 176–204; II, *Trans. AMS* **260** (1980), pp. 223–253.

[3] Barden D., *The Structure of Manifolds*, Ph.D. Thesis, Cambridge Univ., 1963, Cambridge, England.
[4] Bass H., *Algebraic K-Theory*, W. A. Benjamin, New York, 1968.
[5] Bass H., K-Theory and Stable Algebra, *Publ. de l'Inst. des Hautes Etudes Sci.*, **22** (1964).
[6] Bass H., Algebraic K-Theory: A Historical Survey, *Proc. Int. Cong. Math.*, Vancouver, Vol. 1 (1974), pp. 277–283.
[7] Bass H., Heller A. and Swan R., The Whitehead Group of a Polynomial Extension, *Publ. de L'Inst. des Hautes Etudes Sci.*, **22** (1964).
[8] Bass H. and Murthy M. P., Grothendieck Groups and Picard Groups of Abelian Group Rings, *Ann. of Math.* **86** (1968), pp. 16–73.
[9] Browder W., Manifolds with $\pi_1 = Z$, *Bull. AMS* **72** (1966), pp. 238–244.
[10] Browder W., Structure on $M \times R$, *Proc. Camb. Phil. Soc.* **61** (1965), pp. 337–345.
[11] Browder W., Surgery on Simply Connected Manifolds, *Ergebnisse der Math.* **65**, Springer, 1972.
[12] Browder W. and Levine J., Fibering Manifolds Over a Circle, *Comment. Math. Helv.* **40** (1966), pp. 153–160.
[13] Browder W., Levine J. and Livesay R., Finding a Boundary for an Open Manifold, *Amer. J. Math.* **87** (1965), pp. 1017–1028.
[14] Burghelea D., Automorphisms of Manifolds, *Proc. Symp. Pure Math.* **32** (1978) pp. 347–372.
[15] Burghelea D., Some Rational Computations of the Waldhausen Algebraic K-Theory, *Comment. Math. Helv.* **54** (1979), pp. 185–198.
[16] Burghelea D. and Lashof R., Stability of Concordance and the Suspension Homomorphism, *Ann. of Math.* **105** (1977), pp. 449–472.
[17] Burghelea D., Lashof R., and Rothenberg M., *Groups of Automorphisms of Manifolds*, Lecture Notes in Math., Vol. 473, Springer-Verlag, Berlin and NY, 1975.
[18] Cappell S. and Shaneson J., Non-Linear Similarity, *Ann. of Math.* **113** (1981), pp. 315–355.
[19] Carter D., Lower K-Theory of Finite Groups. *Comm. Algebra* **8** (1980), pp. 1927–1937.
[20] Cerf J., La stratification naturelle des espaces de fonctions différentiables réeles et le théorème de la pseudo-isotopie, *Publ. Math. IHES* **39** (1970), pp. 5–173.
[21] Chapman T. A., Topological Invariance of Whitehead Torsion, *Amer. J. of Math.*
[22] Chapman T. A. and Ferry S., Approximating Homotopy Equivalences by Homeomorphisms, *Amer. J. Math.* **101** (1979), pp. 583–607.
[23] Charney R., Homology Stability for GL_n of a Dedekind Domain, *Invent. Math.* **56** (1980), pp. 1–17.
[24] de Rham G., Sur les nouveaux invariants topologique de M. Reidemeister, *Recueil Math.* (T.1) **43** (1935), pp. 737–743.
[25] de Rham G., Sur les complexes avec automorphismes, *Comment. Math. Helv.* **12** (1939), pp. 191–211.
[26] Dwyer W. G., Twisted Homological Stability for General Linear Groups, *Ann. of Math.* **111** (1980), pp. 239–251.
[27] Dwyer W. G., Hsiang W.-C. and Staffeldt R. E., Pseudo-Isotopy and Invariant Theory, I, *Topology* **19** (1980), pp. 367–385.
[28] Farrell F. T., The Obstruction to Fibering a Manifold over a Circle, *Actes du Congrès Int. Math., Nice* (1970), T2, pp. 69–72.

[29] Farrell F. T. and Hsiang W.-C., The Topological-Euclidean Space Form Problem *Invent. Math.* **45** (1978), pp. 181–192.
[30] Farrell F. T. and Hsiang W.-C., On Novikov's Conjecture for Non-Positively Curved Manifolds I, *Ann. of Math.* **113** (1981), pp. 199–209.
[31] Farrell F. T. and Hsiang W.-C., The Whitehead Groups of Poly- (Finite or Cyclic) Groups, *J. London Math. Soc.* **24** (2) (1981), pp. 308–324.
[32] Farrell F. T. and Hsiang W.-C., A Formula for K_1R_a $[T]$, *Proc. Symp. in Pure Math.* **17** (1970), pp. 192–218.
[33] Farrell F. T. and Hsiang W.-C., A Geometric Interpretation of the Künneth Formula for Algebraic K-Theory, *Bull. AMS* **74** (1968), pp. 548–553.
[34] Farrell F. T. and Hsiang W.-C., Topological Characterization of Flat and almost Flat Riemannian Manifolds M^n $(n \neq 3, 4)$ (to appear in *Amer. J. Math.*).
[35] Farrell F. T. and Hsiang W.-C., On the Rational Homotopy Groups of the Diffeomorphism Groups of Discs, Spheres, and Aspherical Manifolds, *Proc. Symp. Pure Math.* **32** (1978), pp. 325–338.
[36] Farrell F. T. and Hsiang W.-C., Manifolds with $\pi_1 = G \times_a T$, *Amer. J. Math.* **95** (1973), pp. 813–848.
[37] Ferry S., Homotoping ε-Maps to Homeomorphisms, *Amer. J. Math.* **101** (1979), pp. 567–582.
[38] Franz W., Über die Torsion einer Überdeckung, *J. Reine Angew. Math.* **173** (1935), pp. 245–254.
[39] Goodwillie T., Ph. D. Thesis, Princeton University, 1982.
[40] Hatcher A. E., Higher Simple Homotopy Theory, *Ann. of Math.* **102** (1975), pp. 101–137.
[41] Hatcher A. E., Concordance Spaces, Higher Simple-Homotopy Theory, and Applications, *Proc. Symp. Pure Math.* **32**, Part I (1978), pp. 3–21.
[42] Hatcher A. E. and Wagoner J., Pseudo-Isotopies of Compact Manifolds, *Astérisque* **6** (1973).
[43] Hsiang W.-C., A Splitting Theorem and the Künneth Formula in Algebraic K-Theory, Algebraic K-Theory and its Geometric Applications, Lecture Notes in Math. **108** (1969), Springer-Verlag, pp. 72–77.
[44] Hsiang W.-C., On π_i Diff M^n, *Geometric Topology* (ed. J. C. Cantrell), Academic Press, 1979, pp. 351–365.
[45] Hsiang W.-C., Manifolds with $\pi_1 = Z^k$, Manifolds — Amsterdam 1970, Lecture Notes in Math. **197** (1971), Springer-Verlag, pp. 36–43.
[46] Hsiang W.-C. and Jahren B., The Rigid Handlebody Theory I, Current Trends in Algebraic Topology, *CMS Conference Proc.*, **2**, Part II (1982), pp. 337–385.
[47] Hsiang W.-C., and Pardon W., When are Topologically Equivalent Orthogonal Transformations Linearly Equivalent? *Invent. Math.* **68** (1982), pp. 275–316.
[48] Hsiang W.-C. and Sharpe R. W., Parametrized Surgery and Isotopy, *Pacific J. Math.*, **67** (1976), pp. 401–459.
[49] Hsiang W.-C. and Staffeldt R. E., A Model for Computing Rational Algebraic K-Theory of Simply Connected Spaces, *Invent. Math.* **68** (1982), pp. 227–239.
[50] Hsiang W.-C. and Staffeldt R. E., *Rational Algebraic K-theory of a Product of Eilenberg–MacLane Spaces*, to appear.
[51] Igusa K., *Higher Singularities of Smooth Functions are Unnecessary*, preprint.
[52] Karoubi M., Fondements de la K-théorie, *Actes du Congrès Int. Math.*, Nice (1970), T. 2, pp. 27–32.

[53] Karoubi M., Théorie de Quillen et homologie du groupe orthogonal; Le théorème fondamental de la K-théorie hermitienne, *Ann. of Math.* **112** (1980), pp. 207–282.
[54] Kervaire M., Le théorème de Barden–Mazur–Stallings, *Comment. Math. Helv.* **40** (1965), pp. 31–42.
[55] Kuiper N. and Robbin J. W., Topological Classification of Linear Endomorphisms, *Invent. Math.* **19** (1973), pp. 83–106.
[56] Kirby R. C. and Siebenmann L. C., Foundational Essays on Topological Manifolds, Smoothings, and Triangulations, *Annals of Math. Studies* **88**, Princeton University Press, 1977.
[57] Lashof R. and Rothenberg M., G-Smoothing Theory, *Proc. Symp. Pure Math. AMS* **32**, Part I (1978), pp. 211–266.
[58] Loday J.-L., K-théorie algébrique et représentations de groupes, *Ann. Sci. Ec. Norm. Sup.* 4e série, **9** (1976), pp. 309–377.
[59] Loday J.-L., *Homotopie des espaces de concordances* [d'après F. Waldhausen], Séminaire Bourbaki, Springer Lecture Notes in Math. 710 (1979), pp. 187–205.
[60] Madsen Ib. and Rothenberg M., Classifying G-Spheres, *Bull. AMS* (New Series) **7** (1982), pp. 223–226.
[61] Mazur B. and Poenaru V., *Séminaire de Topologie combinatoire et différentielle*, Inst. Hautes Etudes Scient. 1962.
[62] Milnor J. W., Two Complexes Which are Homeomorphic but Combinatorially Distinct, *Ann. of Math.* **74** (1960), pp. 575–590.
[63] Milnor J. W., Whitehead Torsion, *Bull. AMS* **72** (1966), pp. 358–426.
[64] Milnor J. W., *Lectures on the h-Cobordism Theorem*, Princeton U. Press, 1965.
[65] Novikov S. P., Pontrjagin Classes, the Fundamental Group, and Some Problems of Stable Algebra. In: *Essay on Topology and Related Topics, Mèmoires dédiés à Georges de Rham*, Springer, 1970, pp. 147–155. Also *Translations Amer. Math. Soc.* **69** (1968), pp. 172–179.
[66] Quillen D., *Higher Algebraic K-Theory*: I, Lecture Notes in Math. 341 (1973), pp. 77–139.
[67] Quillen D., On the Cohomology and K-Theory of the General Linear Group over a Finite Field, *Ann. of Math.* **96** (1972), pp. 552–586.
[68] Quillen D., Higher Algebraic K-Theory, *Proc. Inter. Cong. Math.*, Vancouver, (1974), Vol. 1, pp. 171–176.
[69] Quinn Frank, Ends of Maps I, *Ann. of Math.* **110** (1979), pp. 275–331; II, *Invent. Math.* **68** (1982), pp. 253–424.
[70] Reidemeister K., Homotopieringe und Linsenräume, *Hamburger Abhandl.* **11** (1935), pp. 102–109.
[71] Schultz R., On the Topological Classification of Linear Representations, *Topology* **16** (1977), pp. 263–270.
[72] Segal G., Categories and Cohomology Theories, *Topology* **13** (1974), pp. 293–312.
[73] Siebenmann L., A Total Whitehead Torsion Obstruction, *Comment. Math. Helv.* **45** (1970), pp. 1–48.
[74] Siebenmann L., *The Obstruction to Finding a Boundary for an Open Manifold of Dimension Greater than Five*, Ph.D. Thesis, Princeton Univ., 1965.
[75] Siebenmann L., Deformation of Homeomorphism on Stratified Sets, *Comment. Math. Helv.* **47** (1972), pp. 123–163.
[76] Smale S., On the Structure of Manifolds, *Amer. J. Math.* **84** (1962), pp. 387–399.
[77] Stallings J., On Infinite Processes Leading to Differentiability in the Complement

of a Point, *Differential and Combinatorial Topology,* Princeton Univ. Press, 1965, pp. 245–254.

[78] Steinberger M., *On the Equivalence of the Definitions of the Algebraic K-Theory of a Topological Space,* Lecture Notes in Math. 763 (1979), Springer-Verlag, pp. 317–331.

[79] Steiner R., Infinite Loop Structures on the Algebraic K-Theory of Spaces, *Math. Proc. Camb. Phil. Soc.* **90** (1981), pp. 85–111.

[80] Swan R., *K-Theory of Finite Groups and Orders,* Lecture Notes in Math. 149, Springer-Verlag, 1970.

[81] Swan R., Algebraic K-Theory, *Actes, Congrès Intern. Math.,* Nice, 1970, T. 1, pp. 191–199.

[82] Wagoner J., Delooping Classifying Spaces in Algebraic K-Theory, *Topology* **11** (1972), pp. 349–370.

[83] Waldhausen F., Algebraic K-Theory of Generalized Free Products, *Ann. of Math.* **108** (1978), pp. 135–256.

[84] Waldhausen F., Algebraic K-Theory of Topological Spaces, I, *Proc. Symp. Pure Math.* **32** (1978), pp. 35–60.

[85] Waldhausen F., Algebraic K-Theory of Spaces, a Manifold Approach, *Current Trends in Alg. Topology, CMS Conf. Proc.,* 2, Part I (1982), pp. 141–184.

[86] Wall C. T. C., Finiteness Conditions for CW-Complexes, *Ann. of Math.* **81** (1965), pp. 56–69.

[87] Wall C. T. C., *Surgery on Compact Manifolds,* Academic Press, 1970.

[88] Wall C. T. C., The Topological Space-Form Problems, *Topology of Manifolds,* Markham 1970, pp. 319–331.

[89] Whitehead J. H. C., Simplicial Spaces, Nuclei and m-Spaces, *Proc. London Math. Soc.* **45** (1939), pp. 243–327.

[90] Whitehead J. H. C., Simple Homotopy Type, *Amer. J. Math.* **72** (1950), pp. 1–57.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

PETER D. LAX*

Problems Solved and Unsolved Concerning Linear and Nonlinear Partial Differential Equation

Current research in partial differential equations is extensive, varied and deep. A single lecture, if it is not to be a mere catalogue, can present only a partial list of recent achievements, some comments on the modern style, i.e. the kinds of problems chosen and methods used for solution, and cautious speculations on future trends. The choice of examples is of course shaped by the personal taste of the speaker and limited by his expertise.

The first part of this lecture is such an overview; it is followed by a more detailed discussion of two topics with which the speaker has some familiarity, one concerning a linear, the other a nonlinear problem in partial differential equations.

1a. Linear problems

In the last few years a number of problems concerning linear partial differential operators on manifolds with boundaries have been solved or are nearing solution. Thanks to the researches of Melrose [30], Taylor [37], Ivrii and others we understand well the propagation of signals along reflected, glancing and gliding rays, the clue to many problems in diffraction and scattering. Microlocal analysis, the modern version of wave-ray duality, has provided the tools: pseudo-differential operators, Fourier integral operators, Hamiltonian flows and Lagrange manifolds. In his recent work Charles Fefferman [15] makes use of a sophisticated version of the uncertainty principle. Another versatile modern technique

---

* Preparation of this report was supported by the U.S. Department of Energy under Contract DE-ACO2-76 ERO 3077.

is the use of trace formulas to link spectral and geometric information. The view from scattering theory has also been useful.

For a thorough documentation of the successes of the modern theory of linear partial differential equations we have to await the publication of Hörmander's 3-volume treatise [19], but it is clear that the successes have been so sweeping that they have radically altered the course of research in this field. I believe that in the future we shall see more applications of the methods and results of the theory of linear partial differential equations to other fields of mathematics; examples from the past are of use of PDE methods in several complex variables and quasiconformal mappings. We are likely to see more special questions raised, from sources inside and outside mathematics, and more detailed answers given; mere preoccupation with existence and uniqueness questions is likely to diminish.

Section 2 contains a brief description of wave propagation on complete manifolds of constant negative curvature. One of the tools used is the Radon transform, less popular than its more glamorous sister, the Fourier transform, but more appropriate in some situations. See also [18].

1b. Nonlinear problems

The strides that have been made recently in the theory of nonlinear PDE's are as great as in the linear theory. Unlike the linear case, no wholesale liquidation of broad classes of problems has taken place; rather it is steady progress on old fronts and on some new ones, the complete solution of some special problems, and the discovery of some brand new phenomena. The old tools — variational methods, fixed point theorems, degree of mapping and other topological methods — have been augumented by some new ones. Preeminent for discovering new phenomena, is numerical experimentation; but it is likely that in the future numerical calculations will be part of proofs.

We shall discuss, very briefly, three topics:

(i) Viscous, incompressible flows.
(ii) Hyperbolic systems of conservation laws and shock waves.
(iii) Completely integrable systems.

(i) Viscous, incompressible flows. In spite of a claim by Kaniel, [21], laid to rest by D. Michelson, the existence for all time of strong solutions of the Navier–Stokes equation, and the uniqueness of weak solutions, in three-dimensional space are very much open questions. We have learned more about the singularities of weak solutions, in particular about the Hausdorff dimension of the singular set. Already Leray has shown that every solution is continuous if we eliminate a closed set of t with zero Hausdorff measure of dimension 1/2. B. Mandelbrot has raised the question of what the Hausdorff dimension of the possible singularities of weak solutions is in space and time. The first results on this important question were obtained by V. Scheffer [34]; the latest word is the following theorem of Caffarelli, Kohn and Nirenberg [6]:

The one-dimensional Hausdorff measure of the set of singularities of a suitable weak solution in x, t-space is zero.

Turbulence, surely one of the outstanding problems of mathematics, can be described by the long-time behavior of typical solutions of the NS equations. When viscosity is large compared to the force driving the flow there exists exactly one stationary flow to which all flows tend. As the force is increased, this stationary flow becomes unstable, i.e. any slight perturbation drives it away, perhaps to another, stable, stationary flow. When the force is increased still further, this too becomes unstable and the flow tends to yet another stationary flow or possibly to a periodic flow. As the force is increased further the flow becomes more and more chaotic. This chaotic flow is concentrated around a so-called *attractor set*, i.e. a set consisting of points of accumulation of a single flow driven by a force that is independent of t. Such sets are invariant under the Navier–Stokes flow; concerning these Foias and Temam have proved the following, see [17]:

A bounded set that is invariant under the strong Navier–Stokes flow in a bounded domain has finite Hausdorff dimension.

The dimension of such sets may go to infinity as the viscosity tends to zero.

Further results along these lines have been obtained by P. Constantin and C. Foias.

The simplest testing ground for ideas of instability and turbulence of viscous fluids are the Couette–Taylor flows, i.e. flows between two concentric cylinders, the inner one rotating with some angular velocity ω. If the cylinders have infinite length, then there is a stationary flow that is independent of the angle θ and distance z along the axis of the cylinders. For ω low enough this flow is stable; as ω increases, this flow becomes unstable, yielding stability to another, z-dependent, flow consisting of a stack of Taylor vortices, named after their discoverer. As ω increases further this flow, too, becomes unstable and gives way to a θ-dependent, periodic flow; further increase in ω leads to more and more complicated flows.

There is a wealth of experimental studies of Couette–Taylor flows, revealing a bewildering variety of steady and unsteady flows, see e.g. Benjamin, [3]. The understanding of these (which must also take into account the finiteness of the cylinders) is a profound challenge to theoretical and computational fluid dynamicists.

Flows without any driving force to maintain them decay because the viscous forces dissipate energy. Recently Foias and Saut [16] have shown that the rate of decay is exponential, the same as for the corresponding linearized Stokes flow; they have further shown that the Stokes and the Navier–Stokes flows are linked by a wave operator.

When viscosity is zero, as in the Euler equation, no imposed force is needed to maintain the flow. Existence and uniqueness is known in 2 dimensions but is doubtful when $n = 3$. Extensive calculations by S. Orszag and his colloborators on the Taylor–Green vortex problem, [31], reveal a bewilderingly complicated flow; as time goes on, smaller and smaller scale features appear until the numerical method — a spectral method keeping track of more than 100 million Fourier coefficients — is unable to resolve them. Another set of calculations by Orszag employs the Taylor series in time, up to order 88, summed in a cunning fashion to elude singularities in the complex t-plane; the features revealed in the two calculations are similar.

Another set of calculations, not nearly so machine-intensive as Orszag's has been performed by Chorin, see e.g. [8]. He considers an initial value problem, periodic in space, where vorticity is initially confined to a narrow, slightly crooked tube. The basic variable is vorticity, and the calculation takes into account that the vorticity is confined to the tube, which stretches and twists with the flow. Using a number of bold simplifications the calculation is carried out long enough to indicate that after a finite time the vortex tube will be stretched so thin that its Hausdorff dimension becomes ~ 2.5, a prediction of Mandelbrot's, [29]. Another calculation by Chorin, employing a rescaling reminiscent of the renormalization group of physicists, leads to the same conclusion.

(ii) Hyperbolic systems of conservation laws and shock waves. A system of n conservation laws in one space variable x and in t is of the form

$$u_t + f(u)_x = 0,$$

$u(x, t)$ in $\boldsymbol{R}^n$; the system is strictly hyperbolic if the matrix $\nabla f(u)$ has real eigenvalues for every u in $\boldsymbol{R}^n$.

The basic problem is the initial value problem: given $u(x, 0) = u_0(x)$, show the existence of a solution $u(x, t)$ for all t, in the class of discontinuous functions, satisfying the conservation law in the sense of distributions, and an entropy condition of the form

$$s_t + g_x \leqslant 0,$$

where $s = s(u)$ is an entropy, $g = g(u)$ entropy flux. The two satisfy

$$\nabla s \, \nabla f = \nabla g,$$

and s is required to be a convex function of u.

Numerical evidence indicates strongly that various difference schemes for solving conservation laws converge; yet until recently no proof had been given for systems with more than one state variable u. Similarly, physics strongly suggests but mathematics had been unable to prove that if $u = u_\varepsilon(x, t)$ solves the viscous equation

$$u_t + f(u)_x = \varepsilon D(u)_{xx}, \qquad \varepsilon > 0.$$

$$u(x, 0) = u_0(x),$$

D an appropriate $n \times n$ viscosity matrix, then as ε tends to 0, $u_\varepsilon(x, t)$ converges to a solution of the system of conservation laws that satisfies an entropy condition.

This year Ron Di Perna [12], succeeded in proving such convergence theorems for the equation governing the isentropic flow of a gas satisfying a polytropic equation of state, with the artificial viscosity $D \equiv I$. Among the many ingredients are two beautiful general ideas of Tartar and Murat, [36]. One is a characterization of strong convergence in terms of weak convergence:

Suppose $u_j(y)$ is a sequence of mappings from $\boldsymbol{R}^k$ to $\boldsymbol{R}^m$, uniformly bounded in L^∞; then there is a subsequence such that for every continuous function g in $\boldsymbol{R}^m$ the weak limits

$$\text{w-lim}\, g\big(u_j(y)\big)$$

exists. These limits can be represented as

$$\int g(u)\, d\nu_y(u),$$

where ν_y is a probability measure in $\boldsymbol{R}^m$ parametrized by y in $\boldsymbol{R}^m$. The subsequence u_j converges strongly if the measures v_y have, for each y, a single point as support.

The second idea is compensated compactness: Let v_j and w_j be two sequences of mappings from $\boldsymbol{R}^k$ to $\boldsymbol{R}^k$; if both converge weakly in the L_2 sense to v and w respectively, and if $\operatorname{div} v_j$ and curl w_j lie in compact sets in the H^{-1}_{loc} topology, then

$$\lim \int v_j \cdot w_j dy = \int v \cdot w\, dy.$$

Tartar himself has used these ideas to prove the convergence of viscous solutions for scalar conservation laws; Di Perna has shown how to use them for systems with two variables.

Very little is known about existence of discontinuous solutions in more than one space variable; even short time results are of recent origin, see A. Majda's report to this Congress. Yet numerical calculations, done with care and ingenuity, see e.g. Colella and Woodward, [10], converge and give solutions consistent with experiements.

In one-space dimension we know, at least for simple systems, that shock formation and interaction severely limit the amount of information that a one-dimensional flow field can contain. Something similar must be true in higher dimension, but the mechanism causing it is not understood.

In his report to this Congress, S. Klainerman will describe some recent results on long term existence of regular solutions of the initial value problem for non-linear hyperbolic equations in several space variables.

The question of uniqueness, subject to an entropy condition, is not satisfactorily settled even in one-space dimension, not even for the equation of compressible flow, in spite of the important pioneering work of Oleinik, and more recent work of Di Perna.

There are intriguing open problems concerning stationary transonic flows around given contours, with volocity prescribed at infinity. An ingenious method of Garabedian yields large families of aero-dynamically interesting smooth flows, but a basic theorem of Morawetz shows that for a generic contour no shockless flow exists. The basic problem is to prove the existence of a flow, with shocks, and to prove its uniqueness. Recent numerical calculations of A. Jameson indicate that in the potential flow of approximation there may be many solutions.

(iii) Completely integrable systems. This chapter in mathematics, barely 15 years old, continues to fascinate analysts and algebraists, as well as physicists. The effort has been truly international, and has paid off in the discovery of new completely integrable systems, many of physical interst, some containing two space variables, see Ablowitz and Fokas, [2], and Zakharov's report to this Congress. The algebraic classification of these systems has progressed, see van Moerbeke's report to this Congress, and new connections with other branches of mathematics and physics have been found, such as the τ-function of Sato, Miwa and Jimbo, [33], see Sato's and Takhtajan's reports to this Congress. Three books have appeared recently on solitons and scattering theory [1], [7] and [41], and the work of Beals and Coifman, [4]. The speaker will restrict his remarks to a few scattered comments on the analytic side of the matter.

(a) Solutions of completely integrable partial differential equations lie on infinite-dimensional tori. Numerical experiments with such equations, see e.g. [22], furnish numerical approximations that appear to lie on tori, necessarily finite-dimensional. This indicates that some infinite-dimensional analogue of the KAM theory might be true; no such result is known.

(b) The sine-Gordon equation

$$u_{tt} - u_{xx} + \sin u = 0$$

has explicit solutions

$$u(x, t) = -4\arctan\left[\frac{m}{\sqrt{1-m^2}}\,\frac{\sin\sqrt{(1-m^2}t)}{\cosh mx}\right], \quad m < 1$$

that decay exponentially in x as $|x| \to \infty$ and are periodic in time. If the function $\sin u$ is replaced by $g(u)$, Coron [11] has shown that no such solution can exist when the time period T is $< 2\pi/g'(0)$; Coron and Brezis conjecture that there are no such periodic solutions of any period, except for very special functions g.

(c) The explicit solution of the initial value problem for the KdV equation

$$u_t - 6uu_x + \varepsilon^2 u_{xxx} = 0,$$

$$u(x, 0) = u(x)$$

in terms of the scattering transform makes it possible to determine explicitely the limit of the solution $u(x, t, \varepsilon)$ as $\varepsilon \to 0$. This rather interesting limit is described in Section 3.

These altogether too brief remarks on nonlinear PDE were confined mostly to problems arising in mathematical physics; it is the richest source of such material, but not the only one: geometry is another, see S. T. Yau's report to the Helsinki Congress [39]. The speaker has neither the knowledge nor the time to report on progress in this very active area in the last five years, but he cannot resist mentioning the very recent demonstration by Wente, Steffen, Struwe and Brezis and Coron of the existence of *two* surfaces of constant mean curvature spanning a prescribed plane curve, not too large; the proof is a marvel of subtlety, see Ambrosetti's report to this Congress.

There hasn't even been enough time to mention all the subjects in mathematical physics that have been traditionally, but especially in the recent past, rich sources of problems in nonlinear PDE: elasticity theory, see Ball's report to this Congress, electromagnetic theory and, more recently, magnetohydrodynamics. Two topics which need more help from mathematicians than they are getting now are multiphase flow and combustion, see e.g. [9] and [28]. In both it is of great importance to understand the nature of turbulent regimes; but in multiphase flow, as in aero- and hydrodynamics, turbulence is detrimental; in combustion it is beneficial since it promotes the mixing of fuel and oxydizer. On the other hand shockwaves are detrimental for combustion, since they produce entropy which decreases the efficiency of conversion of heat into mechanical energy.

2. The Laplace–Beltrami operator on complete Riemannian manifolds with constant negative sectional curvature

In a series of papers, [24]–[27], R. S. Phillips and the speaker have analysed, fairly completely, the spectral properties of the Laplace–Beltrami operator on manifolds F as above in the case when F has infinite volume and is geometrically finite. This extends to all dimensions the previous work of Patterson, [32], in the case $n=2$, and allows F to have cusps of all kinds.

The universal cover of F is hyperbolic space H_n; F itself can be identified with the quotient H_n/Γ, Γ a discrete subgroup of isometries of H_n. More concretely F can be identified with a fundamental polyhedron F of H_n mod Γ. Conversely, if Γ is a discrete subgroup of the group of all isometries, then $H_n/\Gamma = F$ is a complete Riemannian manifold with constant negative sectorial curvature; if Γ contains elliptic elements, F has harmless singularities along submanifolds.

We shall use the Poincaré model for $\boldsymbol{H}_n$, i.e. the upper half-space (x, y), x in $\boldsymbol{R}^{n-1}$, $y > 0$, equipped with the metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

The set of points at infinity, $(\beta, 0)$, ∞, is denoted by B.

The L^2 norm and Dirichlet integral, invariant under isometries of $\boldsymbol{H}_u$, will be denoted by $H(u)$ and $D(u)$:

$$H(u) = \int |u^2| \frac{dx\,dy}{y^n}, \qquad D(u) = \int (|u_x|^2 + |u_y|^2) \frac{dx\,dy}{y^{n-2}}.$$

The invariant Laplace–Beltrami operator L_0 is defined in terms of these quadratic forms:

$$H(u, L_0 v) = -D(u, v)$$

for all C_0^∞ functions u and v,

$$L_0 = y^2(\Delta + \partial y^2) - (n-2) y\, \partial y,$$

where $\Delta = \sum \partial x^2$ is the Euclidean Laplace operator. Using the Friedrichs extension, L_0 can be made into a nonpositive self-adjoint operator with respect to H.

Similarly we denote the L^2 and Dirichlet integrals over F by $H_F(u)$ and $D_F(u)$; here u is any C_0^∞ automorphic function with respect to a given discrete subgroup Γ of isometries, and F is a fundamental polyhedron for Γ. In what follows we assume that F has a finite number of sides, i.e. that Γ is *geometrically finite.*

Discrete subgroups can be classified by the geometric properties of their fundamental polyhedra F into the following classes:

(i) F is compact,

(ii) F is noncompact but has finite volume:

$$V(F) = \int_F \frac{dx\,dy}{y^n} < \infty,$$

(iii) F has infinite volume.

The spectral properties of L_0 are sharply different in these cases.

In case (i) it follows by standard elliptic theory that the spectrum of L_0 is standard discrete, i.e. pure point spectrum accumulating only at $-\infty$. The present work is concerned mainly with case (iii); our results are:

(a) L_0 has absolutely continuous spectrum of infinite multiplicity in $\left(-\infty, -\frac{1}{2}(n-1)^2\right)$.

(b) L_0 has at most a finite number of point eigenvalues, all located in the interval $\left(-\frac{1}{2}(n-1)^2, 0\right)$.

(c) L_0 has no singular spectrum; even the point spectrum may be empty. However, Beardon and Sullivan [35] have shown that if F contains a cusp of highest rank, then there is at least one point eigenvalue. We have a new proof of this result, Theorem 6.4 in [26].

Jørgensen, [20] has constructed interesting examples of groups of isometries of $\boldsymbol{H}_3$ whose fundamental polyhedron has infinitely many sides. For these, Epstein, [13], has shown that L_0 has infinite-dimensional spectrum in $(-1, 0)$.

Case (ii) is a curious mixture of (i) and (iii): L_0 has absolutely continuous spectrum in $\left(-\infty, -\frac{1}{2}(n-1)^2\right)$, but only of finite multiplicity, which is equal to the number of cusps. There is no singular spectrum but there may be point spectrum, accumulating at $-\infty$. In many special cases it is known that this point spectrum is ample. In the general case, Selberg has established a relation between the density of the point spectrum and the winding number of the determinant of the scattering matrix; see also pages 205–216 of [24]. To give an absolute estimate of the number of point eigenvalues remains a challenging open problem.

We return now to case (iii). Earlier studies of the continuous spectrum of L_0 proceeded by constructing explicitly a spectral representation of L_0; the generalized eigenfunctions of L_0 entering this spectral representation are Eisenstein series, constructed by analytic continuation.

Our approach is entirely different; it is applicable to representing operators whose continuous spectrum has uniform multiplicity on the whole line. Let A be an *anti-self-adjoint* operator whose spectrum is of uniform multiplicity on the whole imaginary axis. Then the spectral representation for A can be thought of as representing the underlying Hilbert space H by $L^2(\boldsymbol{R}, N)$, N some auxiliary Hilbert space whose dimension equals the multiplicity of the spectrum of A.

Each f in H is represented by an L^2 function $K(\sigma)$, σ in $\boldsymbol{R}$, the values of K lying in N. Since A is anti-self-adjoint, Af is represented by $i\sigma K(\sigma)$.

Denote $U(t)$ the unitary group whose generator is A: $U(t) = \exp tA$. Then

$$U(t)f \leftrightarrow e^{i\sigma t} K(\sigma).$$

The Fourier transform of this representation with respect to σ gives another representation of H by $L^2(\boldsymbol{R}, N)$, where each f in H is represented by

$$f \leftrightarrow k(s), \quad k = \tilde{K}.$$

Then

$$Af \leftrightarrow -\frac{d}{ds} k(s)$$

and

$$U(t)f \leftrightarrow k(s-t), \quad U(t) = \exp tA.$$

This is called *translation representation.*

Of course, conversely, the Fourier transform of a translation representation is a spectral representation.

We show now how to construct a translation representation for the unitary group associated with the non-Euclidean wave equation

$$u_{tt} - Lu = 0,$$

where

$$L = L_0 + \left(\frac{n-1}{2}\right)^2.$$

Note that if (a), (b), (c) hold then, apart of the finite point spectrum, L has continuous spectrum of uniform multiplicity on $(-\infty, 0)$.

The group associated with the wave equation consists of the operators mapping initial data into data at time t:

$$U(t)\colon \{u(0), u_t(0)\} \to \{u(t), u_t(t)\}.$$

The generator of U is

$$A\{u, u_t\} = \{u_t, u_{tt}\} = \{u_t, Lu\} = \{u, u_t\}\begin{Bmatrix} 0 & L \\ 1 & 0 \end{Bmatrix}.$$

Note that $A^2 = \begin{Bmatrix} L & 0 \\ 0 & L \end{Bmatrix}$, thus L having continuous spectrum of uniform multiplicity on $\boldsymbol{R}_-$ is consistent with A having continuous spectrum of uniform multiplicity on $i\boldsymbol{R}$.

Most properties of the non-Euclidean wave equation follow from standard hyperbolic theory:

(i) The initial value problem is properly posed.

(ii) Signals propagate with speed $\leqslant 1$.

(iii) If the initial data are automorphic, so is the solution for all t.

(iv) Energy is conserved, where

$$E = H(u_t) - H(u, Lu) = H(u_t) + D(u) - \left(\frac{n-1}{2}\right)^2 H(u).$$

Finally we have the special property

(v) For n odd the Huygens property holds, i.e. signals propagate with speed $= 1$.

It is not hard to show by integration by parts that for C_0^∞ data in $\boldsymbol{H}_n$, the energy E is positive. We denote by H the completion in the energy norm of C_0^∞ data. It follows from conservation of energy that $U(t)$ is unitary for the energy norm.

We define the *Radon transform* of a function u in $\boldsymbol{H}_n$ by

$$\hat{u} = \int_{\xi(s,\beta)} u\,dS.$$

Here $\xi(s, \beta)$ is the horosphere centered at the point β at ∞, whose distance from the origin is s. It is well known that

$$\widehat{Lu} = e^{-\frac{n-1}{2}s}\, \partial_s^2 e^{\frac{n-1}{2}s}\, \hat{u}.$$

Now take the Radon transform of the wave equation:

$$0 = \hat{u}_{tt} - \widehat{Lu} = \hat{u}_{tt} - e^{-\frac{n-1}{2}s}\, \partial_s^2 e^{\frac{n-1}{2}s}\, \hat{u}.$$

Introduce

$$\hat{u} = e^{\frac{n-1}{2}s} v;$$

then

$$0 = v_{tt} - v_{ss} = (\partial_t + \partial_s)(v_t - v_s),$$

from which it follows that $v_t - v_s$ is a function of $s - t$. We define now

$$R_+\{u, u_t\} = P(v_t - v_s) = P(e^s \hat{u}_t - \partial_s e^s \hat{u}),$$

P an appropriately chosen operator in s that commutes with translation.

R_+ is a translation representation of H for $U(t)$, i.e.

(i) $R_+\{u(t), u_t(t)\}(s) = R_+\{u(0), u_t(0)\}(s-t)$,

(ii) $\int (R_+\{u, u_t\})^2 ds\, dm(\beta) = E(u)$,

(iii) R_+ maps H onto $L_2(\boldsymbol{R}, B)$.

Of course (i) follows from the way R_+ was constructed; for (ii) and (iii), see Theorem 3.4 [25] when $n = 3$ and $P = c\partial_s$.

We turn now to the automorphic case. Here energy, defined as before by

$$E_F(u) = H_F(u_t) + D_F(u) - \left(\frac{n-1}{2}\right)^2 H_F(u),$$

is not necessarily positive. It was shown in Section 4 of [26] that one can add a quadratic function $K(u)$ to the energy so that $G(u) = E_F(u) + K(u)$ is positive, and that K is compact with respect to G. It follows from this that if E_F is negative on a subspace, that subspace is finite-dimensional. Since for $u_t = 0$ the energy is $-H_F(u, Lu)$, it follows that the positive spectrum of L consists of a finite number of eigenvalues. It can be further shown, using the fact that F contains full neighborhoods of points at ∞, that L has no negative eigenvalues, see Theorem 4.8 of [26]; this is a non-Euclidean version of a classical result of Rellich and Vekua.

It follows from the form $E_F = H_F(u_t) - H_F(u, Lu)$ that if u is orthogonal to all the eigenfunctions of L, then $E_F \geqslant 0$. If $u(0)$ and $u_t(0)$ are both orthogonal to all the eigenfunctions it follows that so is $u(t)$ for any t. We denote this space of initial data by H_c. Clearly H_c is invariant under the solution operators $U_F(t)$ for the automorphic solutions of the wave equation.

We define now a translation representation R_+^F of H_c for the group $U_F(t)$; it has $M+1$ components, M being the number of cusps of maximal rank. The zeroth component of R_+^F is R_+, defined as before; each of the remaining components are associated with cusps of maximal rank as follows:

Map the cusp to the point ∞, so that it has the form $F_\infty \times (a, \infty)$, F_∞ a fundamental polyhedron in Euclidean space of the parabolic subgroup keeping ∞ fixed. Since the cusp is of maximal rank, F_∞ has finite volume. Denote by $\bar{u}$ the mean value of u:

$$\bar{u}(g) = \frac{1}{|F_\infty|} \int_{F_\infty} u(x, y)\, dx.$$

Note that the integration is over a part of a horosphere centered at ∞. Now integrate the wave equation over F_∞:

$$\bar{u}_{tt}-y^2\bar{u}_{yy}+(n-2)y\bar{u}_y-\left(\frac{n-1}{2}\right)^2\bar{u} = 0.$$

Introducing $w = y^{(n-1)/2}\,\bar{u}$ and $y = e^s$ as new variables we obtain

$$w_{tt}-w_{ss} = 0.$$

We now define the jth component of the translation representation as w_t-w_s, i.e.

$$R_+^j\{u, u_t\} = c\left(e^{\frac{n+1}{2}s}\,\bar{u}_t-\partial_s e^{\frac{n-1}{2}s}\,\bar{u}\right).$$

In Part I of [27] we show that R_+^F is a partial translation representation of H_c for the group $U_F(t)$; in Part II we prove the completeness of this representation. R_+^F is called the *outgoing* representation. One can define quite analogously the incoming representation R_-^F. The relation of the two is the scattering operator S_F, a notion introduced by Faddeev and Pavlov [14]. As pointed out in Section 4 of [25], the scattering operator is nontrivial already for the case of $\Gamma = \mathrm{id}$, i.e. for the translation representations R_S over all of H_n. This is in sharp contrast to the Euclidean case.

We wish to emphasize that the translation representations are constructed here in purely geometrical terms, i.e. in terms of integrals over horospheres.

3. The zero dispersion limit for the KdV equation

The equation in question is

$$u_t-6uu_x+\varepsilon^2u_{xxx} = 0,$$

and the question under discussion is this: if the initial values of u are fixed,

$$u(x, 0; \varepsilon) = u(x),$$

how does the solution $u(x, t; \varepsilon)$ behave as ε tends to 0?

When we set $\varepsilon = 0$ in the equation, we obtain the reduced equation

$$u_t-6uu_x = 0.$$

This equation has no solution for all t, only in the interval (t_b^{-1}, t_b^+),

$$t_b^- = \left(6 \operatorname*{Min}_x u_x(x)\right)^{-1}, \quad t_b^+ = \left(6 \operatorname*{Max}_x u_x(x)\right)^{-1};$$

here $u(x)$ is the initial value of u. It is reasonable to surmise that for t in (t_b^-, t_b^+), $u(x, t, \varepsilon)$ tends as $\varepsilon \to 0$ to the solution of the reduced equation. What happens when t lies outside this interval? Numerical experiments indicate that over some part of the x-axis, $u(x, t, \varepsilon)$ is oscillatory. As ε tends to 0, the amplitudes of these oscillations remain finite, their wave-lengths are of order ε. Clearly, if we can talk of a limiting behavior as ε tends to 0, this limit can exist only in the weak sense, e.g. in the sense of distributions. This indeed is the case; the speaker and C. D. Levermore have shown, [23], that $u(x, t, \varepsilon)$ tends in the sense of distribution to a limit $\bar{u}$, provided that the initial value $u(x)$ is nonpositive and tends to zero so fast that $\int x u(x)\, dx$ is finite. These papers show not only the existence of a limit but give a fairly explicit formula for the limit $\bar{u}$. For simplicity we take the case when $u(x)$ has a single minimum; then

$$\bar{u}(x, t) = \partial_x^2 Q^*(x, t),$$

where

$$Q^*(x, t) = \operatorname*{Min}_{0 \leqslant \psi \leqslant \varphi} Q(\psi; x, t).$$

Here Q is a quadratic functional of u:

$$Q(\psi;\ x, t) = \frac{4}{\pi}(a, \psi) - \frac{1}{\pi}(L\psi, \psi),$$

L the linear integral operator

$$L\psi(\nu) = \frac{1}{\pi}\int \log\left|\frac{\nu-\mu}{\nu+\mu}\right| \psi(\mu)\, d\mu.$$

The functions admissible in the minimum problem are restricted to lie between 0 and φ, where φ is defined in terms of the initial data as follows:

$$\varphi(\eta) = \operatorname{Re}\int \frac{\eta\, dy}{\left(|u(y)| - \eta^2\right)^{1/2}}, \quad 0 \leqslant \eta.$$

The function a appearing in the linear term in Q depends linearly on x and t, and is defined as follows:

$$a(\eta, x, t) = \eta x - 4\eta^3 t - \theta_+(\eta),$$

where $\theta_+(\eta)$ is a function defined in terms of the inital data. Using the KdV equation it follows that also $u^2(x, t; \varepsilon)$ has a limit $\overline{u^2}$ in the distribution sense, and that

$$\overline{u}_t = \overline{3u_x^2}.$$

Multiplying the KdV equation by u and rewriting the resulting equation as a conservation law shows that

$$\lim_{\varepsilon\to 0} (u^3 + \tfrac{3}{4}\varepsilon^2 u_x^2) = \overline{u^3}$$

exists in the distribution sense, and that

$$\overline{u_t^2} = 4\overline{u_x^3}.$$

Combining this with the explicitly form the $\overline{u}$ leads easily to the formulas

$$\overline{u^2} = \tfrac{1}{3}\partial_x\partial_t Q^*, \qquad \overline{u^3} = \tfrac{1}{12}\partial_t^2 Q^*.$$

The minimum problem defining Q^* is a so-called quadratic programming problem; it turns out that it can be solved explicitly. To see this it is convenient to extend the functions ψ admitted in the minimum problem for all real η as odd functions; as a result we may replace the kernel of the operator L by $\log|\eta-\mu|$. One can show that L is negative definite, and that it is related to the Hilbert transform H as follows:

$$\partial_\eta L = H.$$

We extend now ψ to the upper half of the complex η plane as a harmonic function that vanishes at ∞; ψ can be regarded as the real part of an analytic function of Hardy class. The variational condition for the minimum problem can be regarded as prescribing the real and imaginary parts of the function on complementary subsets of the real axis; for details we refer the reader to [23]. Suffice it here to say that the resulting formulas for $\overline{u}$ show:

(i) For t in (t_b^-, t_b^+), $\overline{u}$ is a solution of the reduced equation, and that in this time interval the convergence of u to $\overline{u}$ takes place not only in the sense of distributions but for each t in the L^2 sense in x

(ii) For t outside the interval (t_b^-, t_b^+), $\bar{u}$ can be described by Whitham's averaged equations, or by the more general equations of Flaschka, Forest and McLaughlin based on multiphase averaging.

(iii) For t tending to ∞, $\bar{u}$ decays like t^{-1}; more precisely

$$\bar{u}(x, t) = -\frac{1}{2\pi}\varphi\left(\frac{1}{2}\sqrt{\frac{x}{t}}\right) + o(t^{-1})$$

for $0 < x/t < 4m$, where $m = \max_x[-u(x)]$. Outside this range $\bar{u}(x, t)$ is $O(t^{-2})$.

The formula for $u(x, t; \varepsilon)$ is obtained from Gardner, Greene, Kruskal and Miura's solution of the KdV equation by the scattering transform. We trace carefully the manner in which this solution depends on ε, and show that as ε tends to zero, it has a limit in the sense of distributions. The nonpositivity of the inital data makes the GGKM solution of the KdV equation particularly simple. A more difficult case has been handled by Venakides [40].

References

[1] Ablowitz M. Jr. and Segur H., *Solitons and the Inverse Scattering Transform*, SIAM Studies in Applied Math., SIAM, Philadelphia, 1981.

[2] Ablowitz M. and Fokas T., *Lecture Notes in Physics*, K. B. Wolf (ed.), Springer-Verlag, 1983.

[3] Benjamin T. B., Bifurcation Phenomena in Steady Flows of a Viscous Fluid, I. Theory, *Proc. R. Soc. London A* **359** (1978), pp. 1–26; II. Experiments, *Proc. R. Soc. London A* **359** (1978), pp. 27–43.

[4] Beals R. and Coifman R. R., Scattering transformations spectrales, et equations d'evolution non linéaires, *Seminaire Goulaoric–Meyer–Schwartz*, 1980–1981, exp. 22, Ecole Polytechnique, Palaiseau.

[5] Brezis H., Periodic Solutions of Nonlinear Vibrating Strings and Duality Principles, *Bull. AMS* **8** (1983), pp. 409–426.

[6] Caffarelli L., Kohn R., and Nirenberg L., Partial Regularity of Suitable Weak Solutions of the Navier–Stokes Equations, *CPAM* **35** (1982), pp. 771–831.

[7] Calagero F. and Degasperis A., Spectral Transform and Solitons, *Studies in Math. and Its Appl.* **13**, North-Holland.

[8] Chorin A. J., The Evolution of a Turbulent Vortex, *Comm. Math. Phys.* (1983), p. 517.

[9] Chorin A. J. *et al.*, *Phil. Trans. Royal Soc. London A* **304** (1982), p. 303.

[10] Colella P. and Woodward P., The Piecewise Parabolic Method for Gas Dynamical Simulation, to appear in *J. Comp. Phys.*

[11] Coron J. M., Period minimale pour une corde vibrante de longeur infinite, *C. R. Acad. Sci. Paris, Ser. A* **294** (1982), pp. 127–129.

[12] Di Perna R., Convergence of the Viscosity Method for Isentropic Gas Dynamics, *Comm. Math. Phys* **91** (1983), pp. 1–30.
[13] Epstein Ch. L., *The Spectral Theory of Geometrically Periodic Hyperbolic 3-Manifolds*, Dissertation GSAS, New York Univ., 1983.
[14] Faddeev L. and Pavlov B. S., Scattering Theory and Automorphic Functions, *Seminar of Steklov Math. Inst. of Leningrad*, **27** (1972), pp. 161–193.
[15] Fefferman Ch. L., The Uncertainty Principle, *AMS Colloquium Lecture Series, Bull. Amer. Math. Soc.* **9** (1983), pp. 129–206.
[16] Foias C. and Saut J.-C., *Asymptotic Properties of Solutions of the Navier–Stokes Equations*, expose au Seminaire de Mathematiques Appliquees du College de France, to appear in 1982, Pitman.
[17] Foias C. and Temam R., Structure of the Set of Stationary Solutions of the Navier–Stokes Equations, *Comm. Pure Appl. Math.* **30** (1977), pp. 149–164.
[18] Helgason S., The Radon Transform, *Progress in Mathematics* **5**, Birkhäuser.
[19] Hörmander L., *Linear Partial Differential Equations*, Springer-Verlag, to appear.
[20] Jørgensen T., Compact 3-Manifolds of Constant Negative Curvature Fibering over the Circle, *Ann. Math* **106** (1977), pp. 61–72.
[21] Kaniel S., *On the Existence of a Global Solution to the Cauchy Problem of Euler and Navier–Stokes Equations*, preprint, University of Jerusalem, 1981.
[22] Lax P. D., Periodic Solutions of the KdV Equation, *CPAM* **28** (1975), pp. 141–188.
[23] Lax P. D. and Levermore C. D., On the Small Dispersion Limit for the KdV Equation, Part I, *CPAM* **36** (1983), pp. 253–290, Part II, pp. 571–593, Part III, pp. 809–830.
[24] Lax P. D. and Phillips R. S., Scattering Theory for Automorphic Functions, *Ann. of Math. Studies* **87**, Princeton University Press, Princeton, N. J., 1976.
[25] Lax P. D. and Phillips R. S., Translation Representations for the Solution of the Non-Euclidean Wave Equation, *Comm. Pure and Appl. Math.* **32** (1979), pp. 617–667.
[26] Lax P. D. and Phillips R. S., The Asymptotic Distribution of Lattice Points in Euclidean and Non-Euclidean Spaces, *Jour. of Functional Analysis* 46 (1982), pp. 280–350.
[27] Lax P. D. and Phillips R. S., Translation Representation for Automorphic Solutions of the Wave Equation in Non-Euclidean Space, Parts I, II and III to appear in *CPAM* **37** (1984).
[28] Majda A., Systems of Conservation Laws in Several Space Variables, *Proceedings* of this Congress.
[29] Mandelbrot B., Intermittent Turbulence and Fractal Kurtosis and the Spectral Exponent $5/3+B$. In: R. Temam (ed.), *Turbulence and Navier–Stokes Equations*, Lecture Notes in Math. **565**, Springer-Verlag, 1976, pp. 121–145.
[30] Melrose R., Transformation of Boundary Problems, *Acta Math.* **147** (1981), pp. 149–235.
[31] Orszag S. *et al.*, *Phys. Rev. Lett.* **44** (1980), pp. 572–575.
[32] Patterson S. J., The Laplace Operator on a Riemann Surface, I, *Compositio Mathematica* **31** (1975), pp. 83–107; II, **32** (1976), pp. 71–112; III, **33** (1976), pp. 227–259.
[33] Sato M., Miwa T., and Jimbo M., *Publ. RIMS, Kyoto Univ.* **14** (1978), p. 223.
[34] Scheffer V., Turbulence and Hausdorff Dimension. In: *Turbulence and the NS Equation*, Lecture Notes in Math **565**, Springer-Verlag, 1976, pp. 94–112.

[35] Sullivan D., Entropy, Hausdorff Measures Old and New, and Limit Sets of Geometrically Finite Kleinian Groups, to appear in *Acta Math.*
[36] Tartar L., Compensated Compactness and Applications to PDE. In: R. J. Knops (ed.), *Research Notes in Mathematics* **39**, Pitman Press, 1979.
[37] Taylor M., *Noncommutative Microlocal Analysis*, preprint, Dept. of Math. SUNY, Stony Brook, NY.
[38] Temam R., *Navier–Stokes Equation and Nonlinear Functional Analysis*, CBMS–NSF Regional Conference Series in Applied Mathematics, 1983, SIAM.
[39] Yau S. T., *Proc. Congress of Mathematicians*, Helsinki, 1978.
[40] Venakides S., *The Zero Dispersion Limit for the KdV Equation*, Dissertation, GSAS, New York University, 1982.
[41] Zakharov V. E., Manakov, Novikov S. V., and Pitayevsky L. P., *Theory of Solitons. The Method of the Inverse Scattering Problem*, Nauka, Moscow, 1980 (in Russian).

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

V. P. MASLOV

Non-Standard Characteristics in Asymptotical Problems

Introduction: Examples of non-standard characteristics. General problems

1. An example of the Klein–Gordon equation. If we asked a physicist what Hamilton–Jacobi equation naturally corresponds to the Klein–Gordon equation

$$h^2\frac{\partial^2 u}{\partial t^2} - h^2\frac{\partial^2 u}{\partial x^2} + m^2c^4u = 0 \tag{0.1}$$

(m, c, h are physical constants), he would write of course the Hamilton–Jacobi equation

$$\left(\frac{\partial S}{\partial t}\right)^2 - \left(\frac{\partial S}{\partial x}\right)^2 - m^2c^4 = 0 \tag{0.2}$$

describing the free motion of a relativistic particle. But the mathematician specializing in hyperbolic equations would write another Hamilton–Jacobi equation,

$$\left(\frac{\partial \Phi}{\partial t}\right)^2 - \left(\frac{\partial \Phi}{\partial x}\right)^2 = 0. \tag{0.3}$$

This equation is a characteristic equation in a standard mathematical sense. Who is right?

The answer is: both the physicist and the mathematician are wrong or, to say it more politely, both of them are right. The nature of the disagreement is easy to see. The physicist is looking for the semi-classical asymptotics of the solution of the Klein–Gordon equation with respect to the parameter h. He sets this parameter to be small.

The mathematician is looking for the asymptotics with respect to "smoothness". Namely, he looks for example for the solution of a problem with singular initial data modulo differentiable functions.

To obtain the characteristic equation (0.3) the mathematician substitutes in the Klein–Gordon equation a solution of the form

$$\theta(\Phi)\varphi_0+\Phi\theta(\Phi)\varphi_1+\dots, \tag{0.3.1}$$

where θ is the Heaviside function, $\Phi \in C^\infty$, $\varphi_i \in C_0^\infty$. Then he puts the coefficient by the main singularity equal to zero.

The physicist obtains the Hamilton–Jacobi equation corresponding to the free motion of a relativistic particle by substituting in the Klein–Gordon equation the function $\varphi\exp(iS/h)$ ($S \in C^\infty$, $\varphi \in C_0^\infty$). Then he puts the coefficient by the main term in the resulting polynomial in h equal to zero.

The problem of constructing the semi-classical asymptotics can easily be reduced to the problem of constructing the asymptotics with respect to smoothness. Namely, consider the solution of the Klein–Gordon equation as a function of one more variable. This variable is our small parameter h. Further, it is convenient to introduce $\lambda = 1/h$. Evidently, the asymptotical expansion of the solution $u(x,t,\lambda)$ of the Klein–Gordon equation is equivalent to the expansion with respect to smoothness of the function $v(x, t, \xi)$, which is the Fourier transform with respect to λ of the function $u(x,t,\lambda)$. And we see, the function $v(x, t, \xi)$ satisfies the hyperbolic equation

$$\frac{\partial^2 v}{\partial t^2}-\frac{\partial^2 v}{\partial x^2}-m^2c^4\frac{\partial^2 v}{\partial \xi^2}=0. \tag{0.4}$$

Indeed, by setting

$$v(x, t, \xi) = (2\pi)^{-1/2}\int e^{-i\lambda\xi}u(x, t, \lambda)\,d\lambda$$

and passing to the Fourier transform in (0.1), we obtain equation (0.4). Note that c in (0.1) and (0.4) may be a smooth function in x.

We search for the parametrix (0.4), i.e., the asymptotic expansion with respect to the smoothness of the solution of equation (0.4) with initial data

$$v|_{t=0} = 0, \qquad \left.\frac{\partial x}{\partial t}\right|_{t=0} = \delta(x-y)\,\delta(\xi-\eta),$$

in the form

$$\int e^{i|p|\Phi(x,\xi,\nu,\omega,t)}\big(\varphi_0(x, \xi, \nu, \omega, t)+|p|^{-1}\varphi_1(x, \xi, \nu, \omega, t)+\dots\big)\,dp,$$

where $p = (p_1, p_2)$ are dual to the variables x, ξ, $|p| = \sqrt{p_1^2+p_2^2}$, $\nu = p_1/|p|$, $\omega = p_2/|p|$, $\Phi, \varphi_0, \varphi_1, \ldots \in C^\infty$.

But the Fourier transform property $|p|$ is considered as a large parameter.

Analogically to the previous procedure, as $|p|\to\infty$ for $\Phi(x, \xi, \nu, \omega, t)$ we obtain the equation

$$\left(\frac{\partial\Phi}{\partial t}\right)^2-\left(\frac{\partial\Phi}{\partial x}\right)^2-m^2c^4\left(\frac{\partial\Phi}{\partial \xi}\right)^2=0. \tag{0.5}$$

By using the expansion of δ-function

$$\delta(x-y)\,\delta(\xi-\eta)=\frac{1}{2\pi}\int\exp\Big(i|p|\big(\nu(x-y)+\omega(\xi-\eta)\big)\Big)\,dp$$

we obtain the condition on Φ:

$$\Phi|_{t=0}=\nu(x-y)+\omega(\xi-\eta).$$

Set

$$\Phi(x, \xi, \nu, \omega, t)=S(x, \xi, \nu, t)+\omega(\xi-\eta),$$

where S is independent of ξ, η (such substitution is possible since equation (0.5) is independent of ξ). Then the equation for S has the form

$$\left(\frac{\partial S}{\partial t}\right)^2-\left(\frac{\partial S}{\partial x}\right)^2-m^2c^4\omega^2=0. \tag{0.6}$$

It is easy to see that for $\omega^2=1$, i.e., when the asymptotics is constructed with respect to the variable ξ only, this equation coincides with (0.2).

Despite such evident correspondence between the semi-classical asymptotics and the asymptotics with respect to smoothness, they were studied independently for a long time. In the middle of the sixties the method of characteristics describing the asymptotics of solutions both with respect to smoothness and to the parameter was expanded upon a wide class of pseudodifferential and h-pseudodifferential operators (i.e., operators with a small parameter h by the derivatives).

Thus there exist two types of characteristics for equations with a small parameter: for the so-called h-differential equations and for more general h-pseudodifferential equations.

The characteristics of the first type describe asymptotical expansions in terms of powers of this parameter. The characteristics of the second type describe asymptotical expansions in terms of smoothness. It is natural to pose for h-pseudodifferential equations the problem of constructing of such compound asymptotical expansions as would include asymptotical expansions both in terms of the parameter and in terms of smoothness. Namely, the problem is to find solutions modulo functions simultaneously smooth and small. For the Klein–Gordon equation the characteristics corresponding to our problem are defined by a one-parametr family of the Hamilton–Jacobi equations (0.6) depending on the parameter $\omega^2 \in [0, 1]$ (such an interval of variation of the parameter ω^2 is the consequence of the equality $\nu^2+\omega^2 = 1$). For $\omega = 0$ these characteristics coincide with the characteristics of the problem of construction asymptotics with respect to smoothness, and for $\omega = 1$ with that of the problem of rapidly oscillating asymptotics.

The compound asymptotics and the family of characteristics corresponding to them play an essential role. For example, the compound asymptotics describe mathematically effects of the Cherenkov type, namely, the phenomenon in with there is a domain of rapid oscillations (light beaming) in the tail of a particle. The particle is described by the δ-function. This domain can be described exactly by a corresponding family of characteristics.

2. Systems of equations of crystal lattice oscillations and difference schemes. Analogous compound asymptotics and corresponding characteristics can be constructed for difference schemes and for systems of a large number of ordinary differential equations.

Consider a simple example of such a system, namely, the atom oscillations in a one-dimensional crystal lattice with the step h on the circle of length $2\pi r_N = Nh$ (N is the number of atoms):

$$\frac{d^2 u_n}{dt^2} = \frac{c^2}{h^2}(u_{n+1}-2u_n+u_{n-1}), \qquad n = 1, \ldots, N. \tag{0.7}$$

Here u_n denotes the deviation of the n-th atom from the equilibrium state, $u_0 = u_N, u_1 = u_{N+1}$.

Assume that the circle radius r_N remains finite where $N \sim 1/h$ is a large number, and let us search for the asymptotical solution under these assumptions.

Consider a smooth $(2\pi r_N)$-periodic function $u(x, t)$ taking the values u_j at the lattice points. Equation (0.7) can be rewritten as follows:

$$\frac{\partial^2 u}{\partial t^2} = \frac{c^2}{h^2}[u(x+h, t)-2u(x, t)+u(x-h, t)].$$

The solution of this equation at the lattice points coincides with the solution of (0.7) and is independent of the initial data outside the lattice points.

Using the identity

$$\exp\left[i\left(-ih\frac{\partial}{\partial x}\right)\right]u(x,t) = u(x+h,t),$$

which can be verified by means of the Taylor expansion or by means of the Fourier transform, we rewrite the latter equation in the form

$$h^2\frac{\partial^2 u}{\partial t^2} + 4c^2\sin^2(-ih\tfrac{1}{2}D)u = 0, \qquad D \equiv \frac{\partial}{\partial x}. \tag{0.8}$$

Even in this simplest case the family of characteristic equations appears not to be standard (Maslov, 1965, [103]). It has the form

$$\left(\frac{\partial S}{\partial t}\right)^2 - 4\frac{c^2}{\omega^2}\sin\left(\frac{\omega}{2}\frac{\partial S}{\partial x}\right) = 0, \tag{0.9}$$

where the parameter ω varies in the interval $[0, 2\pi]$. This fact concerns equation (0.7) but not equation (0.8), and it follows from the fact that the characters of the discrete group, corresponding to the lattice $2\pi k/N$, vary just in this interval.

In the three-dimensional case for the lattice corresponding to an Abelian discrete group the parameters in the characteristic equation vary in the Brillouin zone (well known in the crystal theory [3], p. 100). If the lattice corresponds to a non-Abelian group, then the parameters in the characteristic equation may vary in some extraordinary domains.

Consider the simplest example of a finite-difference equation, namely, the difference scheme

$$h^{-2}(u_n^{m+1} - 2u_n^m + u_n^{m-1}) = c^2h^{-2}(u_{n+1}^m - 2u_n^m + u_{n-1}^m), \tag{0.10}$$

which approximates the wave equation. The family of characteristic equations for (0.10) has the form:

$$\frac{\partial S}{\partial t} \pm 2\omega^{-1}\arcsin\left[c\sin\left(\frac{\omega}{2}\frac{\partial S}{\partial x}\right)\right] = 0,$$

where the parameter ω varies from 0 to 2π. These characteristics define the spread zone of the oscillations for the so-called unity error, i.e., of the oscillation of the solution, which at the initial moment is equal to unity at a point of the net and is equal to zero at other points. First such characteristics for difference schemes were introduced in [106].

For the equations with constant coefficients the spread zone can be calculated directly from the exact solution [146], [39].

3. General setting of the problem of asymptotics compound with respect to smoothness and to the parameter. The problem of constructing asymptotics with respect to smoothness or to the parameter can be regarded as follows: for a given pseudodifferential operator L: $H_s \to H_{s-m}$ in the scale of the Sobolev spaces H_s an almost inverse operator R_N such that

$$LR_N = 1+Q_N \tag{0.11}$$

should be constructed. Here Q_N: $H_s \to H_{s+N}$ is a smoothing operator in the problem of asymptotics with respect to smoothness, or a "small" operator, namely $\|Q_N\|_{H_s \to H_s} = o(h^N)$, in the problems on asymptotics with respect to the parameter h. For the compound asymptotics the operator Q_N is simultaneously a "small" and a smoothing one:

$$\|Q_N\|_{H_s \to H_{s+N}} = O(h^N).$$

When the asymptotics is constructed, the initial equation is reduced to an integral equation of the second kind with a kernel which is not only smooth but also small with respect to the parameter. This fact enables us to prove the existence theorem and to construct estimates of the solution uniform with respect to the parameter. The compound asymptotics can naturally be interpreted as the construction of an almost inverse operator in the scale generated by a pair of commuting operators $A_1 = 1/h$ (the multiplication by the inverse of the parameter) and $A_2 = -i\,\partial/\partial x = -iD$. The Fourier transform with respect to $\lambda = 1/h$ transfers this scale into the usual Sobolev scale.

4. Compound asymptotics with respect to smoothness and the decrease at infinity. More complicated situation arises when we construct asymptotics with respect to an n-tuple of non-commuting operators. For example, for differential equations with growing coefficients, i.e., equations containing both the powers of the differentiation operator $A_2 = iD$ and the powers of the operator $A_1 = x$ it is natural to pose the following problem: an almost inverse operator R_N should be constructed such that the remainder Q_N in (0.11) not only is a smoothing operator but also transforms any function from L_2 into a function rapidly decreasing as x tends to infinity. In this problem the characteristics are defined by an n-tuple of noncommuting operators A_1 and A_2.

Consider the example

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - x^2\big(1+b(x)\big)u, \qquad b \in C_0^\infty(R),$$

The characteristics are defined by means of the following family of non-standard Hamilton–Jacobi equations (Maslov, 1973, [109]):

$$\left(\frac{\partial S}{\partial t}\right)^2 = \left(\frac{\partial S}{\partial x}\right)^2 + \omega\left(1 + b(x)\right),$$

where ω is a parameter varying from zero to one.

For general equations with growing coefficients analogous characteristics are constructed and the global asymptotic of the solution with respect to smootheness and to the growth at infinity is obtained, i.e., an almost inverse operator in the scale induced by the pair A_1, A_2 is constructed (Maslov, 1973, [109]; Maslov and Nazaikinskij, 1979, [117]).

5. The case of degenerate characteristics. In the same way, for equations with singularities or with singular standard characteristics it is sometimes possible to find appropriate self-adjoint operators $A_1, \ldots, A_n$ with respect to which the non-standard characteristics of the equation are non-singular. Then it is possible to construct an almost inverse operator with respect to the scale generated by A_j. The scale is given by the sequence of norms

$$\|u\|_s = \|(1 + A_1^2 + \ldots + A_n^2)^{s/2} u\|.$$

The operator R_N almost inverse to the operator L, satisfies equation (0.1), in which the remainder Q_N is a smoothing operator with respect to the scale:

$$\|Q_N\|_{s \to s+N} < \infty.$$

The operator R_N can be explicitly calculated in terms of functions of non-commuting model operators $A_1, \ldots, A_n$ in the case where they define a representation of a Lie algebra or a nilpotent algebra with nonlinear commutation relations ([112], [64], [66]).

Even in the case of geometrically simplest characteristics, namely, for some degenerating elliptic equations, this idea enables us to construct almost inverse operators in the scale generated by the n-tuple of vector fields $A_1, \ldots, A_n$ which induce a nilpotent Lie algebra (Stein and Folland, 1974, [41]; and others [141], [140], [142], [43], [54]).

6. Example of an asymptotics in the case of characteristics with singularities. We show by means of the simplest example how the ideas discussed above enable us to solve the problem of oscillating solutions for a hyperbolic equation with a singularity in the characteristics.

Consider the problem for the following wave equation:

$$
\begin{gathered}
\frac{\partial^2 u}{\partial t^2} - c^2(x)\,\Delta u = f, \\
u|_{t=0} = 0, \qquad u'_t|_{t=0} = 0, \\
f = \exp\left(\frac{i}{h} S_0(x)\right) \varphi_0(x), \qquad S_0(x) = |x|^2,
\end{gathered}
\tag{0.12}
$$

where $c(x) \in C^\infty(R^n)$, $c(x) \geqslant \delta > 0$, $\varphi_0(x) \in C_0^\infty(R^n)$, $\varphi_0(0) \neq 0$. By the Duhamel principle the solution of this problem is expressed in terms of the solution of the Cauchy problem

$$
\begin{gathered}
\frac{\partial^2 v}{\partial t^2} - c^2(x)\,\Delta v = 0, \\
v|_{t=0} = f, \qquad v'_t|_{t=0} = 0.
\end{gathered}
$$

The standard scheme of constructing the asymptotic solution of the above problem is the following: the solution is presented in the form $\exp(iS/h)\varphi$. Evidently the wave operator acts on the exponent as follows:

$$
\begin{aligned}
e^{-iS/h}&\left[-\left(-ih\frac{\partial}{\partial t}\right)^2 + c^2(-ihD)^2\right] e^{iS/h}\varphi \\
&= \left[-\left(\frac{\partial S}{\partial t}\right)^2 + c^2\left(\frac{\partial S}{\partial x}\right)^2\right]\varphi + \\
&\quad + (-ih)\left[-2\frac{\partial S}{\partial t}\frac{\partial \varphi}{\partial t} + 2c^2\frac{\partial S}{\partial x}\frac{\partial \varphi}{\partial x} - \varphi\frac{\partial^2 S}{\partial t^2} + c^2\varphi\Delta S\right] + \\
&\quad + h^2\left[\frac{\partial^2 \varphi}{\partial t^2} - c^2\Delta\varphi\right].
\end{aligned}
$$

Hence $\partial S_0/\partial x = 0$ for $x = 0$, and the solution of the characteristics equation

$$
-\left(\frac{\partial S}{\partial t}\right)^2 + c^2\left(\frac{\partial S}{\partial x}\right)^2 = 0, \qquad S|_{t=0} = S_0,
$$

is a non-smooth function at this point. Thus the standard scheme of constructing the asymptotics with respect to the parameter cannot be applied to this case.

It is evident that the characteristics of the wave equation corresponding to the asymptotics with respect to smoothness (parametrix) (0.11) have no singularities. Indeed, the equation is homogeneous in operators iD; hence the bicharacteristics start from the sphere $|p| = 1$ and the singular point $p = 0$ drops out in this case. The asymptotics with respect to smoothness enables us to represent the solution (0.12) in the form $u = (R_N + r_N)f$ where R_N: $H_s \to H_{s+1}$ is an explicitly calculated operator, and r_N: $H_s \to H_{s+N}$ is a smoothing operator. The operator r_N in this example transfers the right side to a function whose norm in the space H_{s+N} is of order $h^{n/2}$, since this function is a convolution of the exponent $\exp(iS/h)$ with the (smooth) kernel of the operator r_N. The other terms of the asymptotics of function $r_N f$ with respect to h are obtained by means of the solution of the wave equation with zero Cauchy data and smooth right sides. This problem differs from the initial one in that the wave operator should be inverted on a function uniformly smooth (non-oscillating) with respect to the small parameter. Such a solution can easily be obtained by means of computer.

The accuracy of the approximation of the solution by the leading term of the asymptotics depends on the type of the initial conditions. For example, if $S_0(x) = |x|^4$ then the leading term $R_N f$ approximates the solution modulo $O(h^{n/4})$. Thus the estimate of the leading asymptotical term is connected with the individual initial condition. The leading term $R_N f$ can asymptotically be represented in the form of an integral whose integrand oscillates in the parameter h and has a singularity. For example in case $c = 1$, $n = 3$ we have

$$R_N f = \frac{1}{4\pi} \int\limits_{|x-\xi| \leqslant t} |x-\xi|^{-1} \exp(i|\xi|^2/h)\varphi_0(\xi)\, d\xi.$$

The leading term of the asymptotics in the case of variable coefficients and an arbitrary dimension has the same properties.

This scheme is applicable to the case of a general operator equation with characteristics non-singular with respect to an n-tuple of operators and singular with respect to another n-tuple of operators.

Apparently, this scheme covers the results of Guillemin, Sternberg, Uhlmann and others [37], [43], [44], [49], [50], [127], [5].

7. General setting of the problem of characteristics. We have considered a number of simple examples which show that it is necessary to give a general definition of characteristics. The main part of this general notion of characteristics is the possibility of connecting different areas in physics and mathematics. Then the methods of one area can be applied in another. For example, it turns out that the following problems, apparently unconnected, are of the same mathematical nature: the behaviour of the solutions of the equations of the dynamics of a viscous liquid far ahead of the shock wave and large deviations in the theory of probabilities; the Cherenkov effect and the diffusion of unity error in a difference scheme; effects of dissociation of molecules and conical refraction in acoustics and so on. These effects often involve difficulties, which arise in proving the existence theorems for pseudodifferential equations.

First of all, in order to define characteristics, we need an n-tuple of operators, namely, model operators with respect to which the asymptotics is constructed. Further, the original equation should be expressed in terms of those model operators. This procedure itself demands preliminary investigation, namely, the construction of the calculus of non-commuting operators.

In this paper I have no opportunity to give a general definition, since it demands preliminary considerations. I prefer to show, by means of simple but specific examples, the basic ideas which enable us to construct the characteristics in a more general situation.

Part I. The non-standard characteristics of linear equations and equation with non-local nonlinearity

1. Pseudodifferential operators with symbols and characteristics on arbitrary symplectic manifolds. Quantization conditions for coordinate-momenta. Consider the first generalization of the notion of characteristics. We shall describe below the difficulty which arises in asymptotical problems for asymptotics in terms of the parameter. It also arises for compound asymptotics in terms of an n-tuple of operators. The phase space for these problems is not necessarily a cotangent bundle of manifolds.

Even in the simplest example of the equations of oscillations of atoms which we mentioned above the phase space is not a cotangent bundle. In fact, the configuration space in this problem is a two-parameter family of circles. This family depends on the continuous parameter h tending to 0 and on the discrete parameter N, which tends to infinity. Note that the length of the circle is equal to Nh. The character arguments of the discrete group of shifts of the lattice vary from 0 to 2π, and the space of momenta is the unit radius circle. Thus, in this case the phase space is a family of two-dimensional tori. The surface of the torus is obviously equal to $2\pi Nh$; hence the following equality holds:

$$\frac{1}{2\pi h}\int dp\wedge dxN. = \tag{1.1}$$

Consider now a one-parameter family of tori. If the left-side expression in the equality (1.1) is an integer, then it is possible to construct a calculus of pseudodifferential operators; namely, with any symbol f which is a smooth function on the torus one can associate an operator $\hat{f}$ which is defined on functions on the circle $S^1 = 0 \leqslant x \leqslant 2\pi r_N$, so that the following formulas hold: the commutation formula

$$[\hat{f}, \hat{g}] = -ih\widehat{\{f, g\}} + O(h^2) \tag{1.2}$$

and the formula of the operator action on the exponent

$$\begin{aligned} e^{-iS/h}\hat{f}(e^{iS/h}\varphi) &= f(x, dS(x))\varphi(x) + \\ &+(-ih)\left\{\frac{\partial f}{\partial p}(x, dS)\frac{\partial \varphi}{\partial x} + \frac{\varphi}{2}\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial p}(x, dS)\right) - \right. \\ &\left. -\frac{\varphi}{2}\operatorname{tr}\frac{\partial^2 f}{\partial x\,\partial p}(x, dS)\right\} + O(h^2). \end{aligned} \tag{1.3}$$

Here curly brackets denote the Poisson brackets, φ is an arbitrary smooth function on the circle, S is a real function on the circle which defines a smooth curve $p = dS(x)$ on the torus, and the estimate $O(h^2)$ in (1.3) is in the norm L_2.

As in the case of the Euclidean space the operators $\hat{f}$ will be called h-pseudodifferential operators (with symbols on the torus). We shall use the notation

$$\hat{f} = f\left(x, \frac{h}{i}D\right).$$

Now to a symbol on the torus $f = c^2\sin^2(p/2)$ there corresponds an h-pseudodifferential operator of the form

$$\hat{f} = c^2\sin^2\left(\frac{h}{i}D\right).$$

Consider an equation with this operator

$$h^2\frac{\partial^2 u}{\partial t^2} + 4c^2\sin^2\left(\frac{h}{i}D\right)u = 0, \qquad u|_{x=2\pi r_N} = u|_{x=0}. \tag{1.4}$$

Denote by u_k the value of the function u at the point $x = kh$. Then we can write

$$u_{k+1} - 2u_k + u_{k-1} = (e^{h\partial/\partial x} - 2 + e^{-h\partial/\partial x})u|_{x=kh}$$
$$= -4\sin^2\left(-\frac{ih}{2}\frac{\partial}{\partial x}\right)u\Big|_{x=kh}.$$

Thus equation (0.7), describing the oscillations of the atom lattice on a circle, is equivalent to the h-pseudodifferential equation (1.4), whose phase space is a torus satisfying condition (1.1).

Now consider the general case: the family of symplectic manifolds depending on a parameter μ, varying in a compact. It is convenient to consider that the manifold is fixed, and that the symplectic structure on it depends on the parameter μ (i.e., it is a closed non-degenerate 2-form $\Omega^{(\mu)}$ or the Poisson bracket $\{\ldots,\ldots\}^{(\mu)}$). We assume that for $\mu = \mu(h)$ the following condition holds: over any two-dimensional cycle the integral of the symplectic form $\Omega^{(\mu(h))}$ divided by $2\pi h$ coincides (modulo integer numbers) with half of the value of the second Stiefel–Whitney class, i.e.,

$$\frac{1}{\pi h}[\Omega^{(\mu(h))}] = W_2(\mathrm{mod}\,2Z). \tag{1.5}$$

Then with each smooth function f on the phase manifold one can associate an operator $\hat{f}$, so that the commutation formula (1.2) (with the Poisson bracket $\{\ldots,\ldots\}^{(\mu(h))}$) holds and formula (1.3) holds locally (Karasev and Maslov, 1981, [66], details in [67], [68], reproduced in [120], [124]).

Call equation (1.5) the quantization condition for coordinate-momenta. The value $\frac{1}{2}W_2$ in (1.2) will be called the vacuum correction. It is remarkable that condition (1.5) (with the zero vacuum correction $W_2 = 0$) arose already in the construction of a rather narrow class of pseudodifferential operators with locally linear in ϱ symbols, namely, the differential operators of the first order (Kostant, 1970, [77]; Souriau, 1966–1970, [149–151]). For the half-integer vacuum correction (there are important examples, see e.g. [42]) within the framework of the first order operators, condition (1.5) was obtained for the Kaehler manifolds (Czyż, 1979, [23]) and in the case of general real manifolds (Hess, 1981, [55]).

By constructing the calculus of pseudodifferential operators on the orbits of a compact Lie group condition (1.5) numerates the irreducible representations of the group. Thus, the quantization of coordinate-momenta on the orbits coincides with the Weyl rule of integer major weights of irreducible representations (Borel and Hirzebruch, 1959, [16]; Kirillov, 1968, [74]). The vacuum correction is then equal to zero.

It should be noted that in general phase manifolds the quantization condition for coordinate-momenta is a sufficient condition for the existence of the canonical operator on Lagrangian submanifolds [68]. This fact enables us to apply to general phase manifolds the theory of global asymptotical solutions of h-pseudodifferential equations, which is constructed in detail in R^{2n} [103]. The bicharacteristics of such h-pseudodifferential operators belong to those phase manifolds on which the operator symbols are given. The global calculus of h-pseudodifferential operators can also be defined on symplectic V-manifolds [67].

Note that in constructing the calculus of ordinary pseudodifferential operators on an important class of homogeneous symplectic manifolds (Boutet de Monvel and Guillemin, 1981, [18]) the quantization conditions do not arise, see [68].

2. Electron terms. Now we consider the second generalization of the notion of characteristics. Recall that the characteristics equations for a hyperbolic system of equations of the first order are obtained from the characteristic matrix by equating its determinant to zero. In the same way we obtain the characteristics for the systems of h-pseudodifferential equations

$$ih\frac{\partial u}{\partial t} + L\left(x, \frac{h}{i} D\right) u = 0, \qquad x \in R^n. \tag{1.5.1}$$

Here $u = (u_1, \ldots, u_m)$ and the symbol $L(x, p)$ denotes a smooth matrix-valued function, diagonalized by a smooth transformation. The equations of characteristics for system (1.5.1) are defined by the eigenvalues $\lambda_j(x, p)$, $j = 1, \ldots, l$ ($l \leqslant m$) of the matrix $L(x, p)$ and have the form

$$\frac{\partial S}{\partial t} + \lambda_j\left(x, \frac{\partial S}{\partial x}\right) = 0, \qquad j = 1, \ldots, l. \tag{1.5.2}$$

It is natural to pass from the finite-dimensional space, where the symbol $L(x, p)$ acts to an infinite-dimensional space, which results in equations with the operator-valued symbol $\hat{L}(x, p)$. If the spectrum of the operator $\hat{L}(x, p)$ is discrete and its eigenvalues have constant multiplicity, the asymptotical solution is constructed as in the finite-dimensional case (Maslov, 1965, [103], [104]). In this case the equations of characteristics are defined in the same way as the equations of characteristics for systems. They have the form of equations (1.5.2), but the functions $\lambda_j(x, p)$ are now the eigenvalues of the operator $\hat{L}(x, p)$. *If the spectrum of the operator $\hat{L}(x, p)$ is continuous, then the equations of characteristics are defined by the poles of the analytical continuation of its resolvent which are closest to the real axis.* This fact is used in the collision theory and in the theory of the decay of nuclei (the Gamov theory). Since the poles are complex, the characteristics are also complex in this case.

The eigenvalues $\lambda_j(x, p)$ of the symbol of the h-pseudodifferential operator are called terms or effective Hamiltonians in physical literature.

As an example consider the problem of interaction of heavy and light particles (the nuclei and the electrons). This problem is studied in the quantum theory of molecules and in the theory of collisions. In this case the small parameter appears in the Schrödinger equation only at the derivative corresponding to heavy particles. In the simplest one-dimensional model the equation is

$$ih\frac{\partial \psi}{\partial t} = -h^2 \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial y^2} + V(x, y)\psi.$$

The operator-valued symbol has the form

$$\hat{L}(x, p) = p^2 - \frac{\partial^2}{\partial y^2} + V(x, y).$$

Its eigenvalues (the electron terms) define the characteristics which describe the motion of heavy particles along the classical paths in the field generated by the light quantum particles.

The concept of equations with operator-valued symbols has a general character and can be used, in particular, to obtain the characteristics and the corresponding asymptotical expansions in terms of "smoothness". Examples of such characteristics are given by Grushin, 1972, [46], [47], and, in fact, by Boutet de Monvel and Guillemin, 1981, [18], and by Guillemin and Sternberg, 1979, [49].

It should be noted, that this general concept allows us to connect problems which seem, on the surface, absolutely different. For example, physicists' papers on the problem of predissociation of molecules with the intersection of electron terms helped Kucherenko, 1974, [84] in constructing the parametrix for non-strictly hyperbolic equations.

3. Pseudodifferential operators with complex characteristics. Global asymptotics. Now consider the third generalization of the notion of characteristics. Difference schemes and Markovian chains can now be represented as h-pseudodifferential equations. Then it is possible to obtain their characteristics. Note that the equations of characteristics are complex for Markovian chains, non-symmetric difference schemes, equations of the principal type, and also for the problems of the decay of nuclei mentioned above. For example, the simplest difference scheme $u_i^{n+1}-u_i^n = \frac{\tau}{h}(u_{i+1}^n-u_i^n)$ on the circle (τ is the step in t, and h is the step on the circle), represented in the form of an h-pseudodifferential operator

$$[(e^{\alpha h\partial/\partial t}-1)-\alpha(e^{h\partial/\partial x}-1)]u = 0, \qquad \alpha = \frac{\tau}{h},$$

leads directly to the complex characteristics equation

$$\frac{\partial S}{\partial t}+H\left(\omega, \omega\frac{\partial S}{\partial x}\right) = 0,$$

where $\omega \in [0, 2\pi]$ is the parameter (see equation (0.9)), and

$$H(\omega, p) = \frac{1}{i\omega\alpha}\ln\left(1+\alpha[\exp(ip)-1]\right)$$

is a complex-valued Hamilton function *with the non-positive imaginary part.* The equations describing processes with absorption lead to complex characteristic equations with analogical properties.

Complex solutions of real analytical equations of characteristics were considered in the famous paper by Leray, Gårding and Kotake, 1964, [96], in connection with the study of the singularities of non-analytical solutions of partial differential equations. In physical literature complex solutions of real analytical Hamilton–Jacobi equations and of the Hamilton system were used long ago (Keller, 1956, in the problem of reflection [70]; Maslov, 1963, in the problem of scattering [102]; Kravzov, 1967, in the analogous problem of refraction [81], and others). This approach encountered essential difficulties which arise in obtaining analytical solutions of the Hamilton–Jacobi equation and choosing the right branch of the multi-valued solution. These difficulties were avoided when the problem was solved modulo $O(h^N)$ (or modulo n times differentiable functions) with the help of constructions based on the following simple idea. Let the asymptotical solution have the form $\exp\left(\frac{i}{h}S(x)\right)\varphi(x)$, where $\operatorname{Im} S \geqslant 0$ (which is necessary for the boundedness of the solution as $h\to+0$). It is clear that the values of the functions S and φ in the domain $\operatorname{Im} S \geqslant \delta > 0$ are not essential, since the solution in this domain vanishes with the accuracy considered. Then the imaginary part of the function S acts as an additional small parameter which follows from the estimate:

$$(\operatorname{Im} S)^{\nu} \exp(ih^{-1}S) = O(h^{\nu}).$$

Thus one can construct asymptotical analogues of the analytic Hamiltonian formalism, in which *analyticity is reduced by almost analyticity*, namely, it is required that *the Cauchy–Riemann conditions should be satisfied* mod $((\operatorname{Im} S)^N)$.

This idea was used in [108] heuristically in the theory of the complex germ. Almost analytical formalism is also based directly on this idea. It was also used by Treves in [154] to construct the parametrix of the equation of the principal type.

Since 1965, [103], the n-dimensional submanifolds of the phase space $R_x^n \times R_p^n$, which annihilate the symplectic form, have been the basic geometrical construction in the theory of characteristics. The author called them Lagrangian manifolds.

The geometrical basis of almost analytical formalism is the notion of the analytical Lagrangian manifold. It is locally a $2n$-dimensional real submanifold in the complex phase space $C^{2n} = R^{4n}$. This approach is close to the analytical theory and has theoretical harmony.

The main geometrical construction in another approach is the "Lagrangian manifold with the complex germ". The structure of the complex germ is given on an n-dimensional real manifold Λ which is "almost Lagrangian" by means of imbedding into the complex phase space $C_Q^n \times C_p^n$. This means that on Λ a non-negative function D (called the dissipation) and the function W satisfying the condition $dW = PdQ + O(D)$ are given.

The dissipation condition is assumed to be satisfied. This means that the planes tangent to Λ are C-Lagrangian on the zero set of dissipation; namely, they are real-similar, they annihilate the form $dP \wedge dQ$ and satisfy the positivity condition $\operatorname{Im}(P_\alpha g, Q_\alpha g) \geqslant 0$, $\forall g \in C^n$, where $\alpha = (\alpha_1, \ldots, \alpha_n)$ are the coordinates on the plane.

The advantage of the theory of the complex germ over the almost analytical theory is the following. Asymptotical formulas obtained by means of the complex germ are more constructive and simpler. For example, for the construction of the global asymptotics of an h-pseudodifferential equation of the first order it is sufficient to solve only the Hamilton system and the system in variations.

Namely, let $H(x, p)$ be a smooth complex-valued Hamilton function corresponding to the asymptotical problem, $\mathscr{H} = \operatorname{Re} H$, $\tilde{H} = \operatorname{Im} H \leqslant 0$.

To construct the asymptotics in the theory of the complex germ the (real) Hamilton system and the system of equations in variations should be solved:

$$\dot{q} = \mathscr{H}_p, \quad \dot{p} = -\mathscr{H}_q,$$

$$\dot{z} = \mathscr{H}_{qq}z + \mathscr{H}_{qp}w + i\tilde{H}_q,$$

$$\dot{w} = -\mathscr{H}_{pq}z - \mathscr{H}_{pp}w - i\tilde{H}_p.$$

Almost analytical formalism and the theory of the complex germ enable us to construct global asymptotical solutions of the problems mentioned above and to obtain simpler formulas by using additionally only the solution of system in variations. The last fact is essential for constructing the asymptotics of solutions of stationary problems (Maslov, 1977, [113]). Thus the asymptotics of some concrete quantum-mechanical spectrum were obtained [113], [32].

Analogous spectral series were obtained by the method of model problems for the Laplace equation on a Riemann manifold (see Babich and Buldirev, 1972, [6]; Lazutkin, 1969, [92]).

The theory of the complex germ permitted the construction of a complete system of semi-classical coherent states for the Schrödinger operator, the Klein–Gordon operator, and the Dirac operator (Bagrov, Belov, and Ternov, 1982, [8], [7]).

I call asymptotic formulae obtained by means of the theory of the complex germ additive asymptotical formulae. They have the form $\psi = \psi_0 + O(h^N)$, where ψ is the exact solution and ψ_0 is the asymptotical one. Such asymptotical formulae in the diffraction theory describe only the domain of partial shadow. In the theory of probabilities they describe only normal deviations.

Distinct from additive asymptotical formulae, multiplicative formulae have the form $\psi = \psi_0(1 + O(h^N))$ and describe in the diffraction theory the whole domain of shadow. In the theory of probabilities they describe large deviations.

The difference between these two situations can be shown by means of the following trivial example: the multiplicative asymptotical formula of the function $\exp\{(-x^2-x^4)/h\}$ is the function itself but the additive asymptotical formula for this function has the form $\exp\{-x^2/h\} + O(h)$ as the following condition holds:

$$\operatorname*{Max}_x (x^4 \exp\{-x^2/h\}) = (2h/e)^2, \quad \text{and so on}.$$

4. Problems of the logarithmic asymptotics of the solution. The class of equations of the tunnel type. Instanton as the logarithmic limit.

4.1. *Equations of the tunnel type.* Now we consider the fourth generalization of the notion of characteristics. Equations whose solutions do not oscillate at all, but only damp, are of great importance. We call such equations tunnel equations. The asymptotical formulas for the solutions of such equations are also constructed with the help of characteristics, but the latter are obtained in another way, which is the result of the general definition of characteristics.

Compare the characteristics of equations with rapidly oscillating solutions and those of equations with rapidly damping solutions. For this purpose we consider the Schrödinger equation and a parabolic equation. They are very similar. On the left we write the Schrödinger equation with a small parameter h and its solution, on the right we write the parabolic equation, which also has a small parameter h.

Schrödinger equation

$$ih\frac{\partial\psi}{\partial t} = -h^2\frac{\partial^2\psi}{\partial x^2} + V(x)\psi.$$

The Green function for $V(x) = 0$ has the form

$$\psi_0 = \frac{1}{\sqrt{4\pi ith}}\exp\left[\frac{i(x-\xi)^2}{4th}\right].$$

Parabolic equation

$$h\frac{\partial u}{\partial t} = h^2\frac{\partial^2 u}{\partial x^2} - V(x)u.$$

The Green function for $V(x) = 0$ has the form

$$u_0 = \frac{1}{\sqrt{4\pi th}}\exp\left[-\frac{(x-\xi)^2}{4th}\right].$$

As $h\to 0$, the function ψ_0 rapidly oscillates and the function u_0 rapidly damps without oscillations. For $V(x) \in C_0^\infty$ and for small t the asymptotics of the Green function for the Schrödinger equation has the form:

$$\psi = \frac{1}{\sqrt{2\pi ih}}\sqrt{\left|\frac{\partial^2 S(x, \xi, t)}{\partial x\,\partial\xi}\right|}\exp\left[-\frac{i}{h}S(x, \xi, t)\right] + O(h),$$

where the function $S(x, \xi, t)$ satisfies the h-characteristic equation

$$\frac{\partial S}{\partial t} - \left(\frac{\partial S}{\partial x}\right)^2 - V(x) = 0.$$

The asymptotics of the Green function for the parabolic equation as $h \to 0$ has the form

$$u = \frac{1}{\sqrt{2\pi h}} \sqrt{\left| \frac{\partial^2 S_1(x, \xi, t)}{\partial x \, \partial \xi} \right|} \exp\left[-\frac{S_1(x, \xi, t)}{h} \right] \left(1 + O(h)\right),$$

where the function $S_1(x, \xi, t)$ satisfies the Hamilton–Jacobi equation

$$\frac{\partial S_1}{\partial t} + \left(\frac{\partial S_1}{\partial x} \right)^2 - V(x) = 0.$$

It is natural to consider the last equation as the equation of characteristics for the asymptotical problems corresponding to the parabolic equation.

Note the following. The Hamilton–Jacobi equation corresponding to the Schrödinger equation has a physical sense. Moreover, for the problem which is described by the Schrödinger equation with a small parameter h this equation was considered earlier, since it describes the classical motion of a particle. The physical sense of the Hamilton–Jacobi equation is not clear, but the solution u of the parabolic equation has a remarkable property: *there exists a limit*

$$\lim_{h \to 0} h \ln u = -S_1(x, \xi, t).$$

The global asymptotics of the parabolic equation in the multidimensional case was constructed by Maslov in 1965, [103]. For an important class of equations which are connected with the probability problems the logarithmic limits of the solutions were first obtained in the outstanding papers of Varadhan, 1966, [157]; 1967, [158]; and Borovkov, 1967, [17].

The global asymptotics for the heat conduction equation with varying coefficients in degenerated case was constructed by Molchanov, 1975, [130], Kifer, 1976, [73], Maslov and Chebotarev [20].

General systems of equations with this property were published by the present author as late as 1981, [121], though the term "tunnel equations" had been introduced in 1965 in "The perturbations theory and asymptotical methods". In the preface to that book the author promised to devote his next work to these equations. However, he has not done so. He has abandoned this theme, considering the class of tunnel equations to be too narrow. It is difficult to invent examples of systems of equations which would belong to this class, besides the parabolic equations, which are the only equations in the class of differential equations containing the derivative of the first order with respect to time. Then it turned out that the narrowness is not a defect of this class but on the contrary, its advantage, since it absorbs, from among all the possible

Hamiltonians, the right physical models. Probably, the only h-pseudo-differential (not differential) equation of the first order with respect to time which belongs to this class is the general integro-differential Kolmogorov equation in the theory of probabilities.

We give the tunnel conditions in the case of systems of h-pseudo-differential equations of the form

$$h\frac{\partial u}{\partial t} = A\left(-h\frac{\partial}{\partial x}, x, t\right)u. \tag{1.6}$$

Here $A(p, x, t)$ is a $(2n+1)$-parametric $m\times m$ matrix-valued function whose elements are entire functions of the argument p and smooth functions in x and t. Denote by $H_a(p, x, t)$, $a \leqslant m$ the eigenvalues (Hamiltonians) of the matrix $A(p, x, t)$.

DEFINITION. A system is called a *system of the tunnel type* if all its Hamiltonians H_a have constant multiplicity as $p \in R^n \setminus \{0\}$, $t\in [0, T]$ and satisfy the following conditions:

(1) for $t \in [0, T]$, $x \in R^n_x$, $p \in C^n \setminus \{\bigcap_{i=1}^{n} \operatorname{Re} p_i = 0\}$ the function $H(p, x, t)$ depends regularly on its argument p,

(2) $\max_\eta \operatorname{Re} H(p+i\eta, x, t) = H(p, x, t)$, $|p| \neq 0$,

(3) the Lagrangian L is equal to $\langle p, H_p(p, x, t)\rangle - H(p, x, t) \geqslant 0$,

(4) $\det\|H_{pp}(p, x, t)\| \neq 0$ for $|p| < \infty$.

These conditions are valid for the linearized Navier–Stokes system with small viscosity, for systems of equations arising in the theory of plasma, in magnetic hydrodynamics (e.g., see [87]), in the theory of elasticity, for the Boltzmann equation, for systems of the Semenov equations (e.g., see [145]), of the chain reaction, for the equation of drop coagulation (e.g., see [71]).

As an example we consider the equation which describes the Foycht model in the theory of elasticity:

$$\frac{\partial^2 u}{\partial t^2} = E\frac{\partial^2 u}{\partial x^2} + h\frac{\partial^3 u}{\partial x^2\,\partial t},$$

where $E = E(x) \geqslant 0$ is a smooth function, and $h > 0$ is a small parameter. To this system correspond the following Hamiltonians:

$$H_{1,2} = \frac{p^2}{2} \pm p\left(\frac{p^2}{4} + E\right)^{1/2}.$$

The verification of the tunnel conditions for these Hamiltonians requires hard calculations.

4.2. *Construction of the asymptotics of the solution at focal points.* The construction of the asymptotics of the Green function near the focal points for the equation of the tunnel type (1.6) is analogous to the corresponding construction for the oscillating solutions of h-pseudodifferential equations.

Consider this analogy in the case of the scalar equation $ih\,\partial/\partial t = H(-ih\,\partial/\partial x, x, t)\psi$. For simplicity, let $\mathrm{tr} H_{px} = 0$. In this case the asymptotical Green function is constructed with the help of the canonical operator[1] on the Lagrangian graph of the canonical transformation g_H^t, which corresponds to the shift along the trajectories of the Hamiltonian H through time t. The canonical operator defines the projective representation K of the group of canonical transformations of the phase space R^{2n} in the space of functions asymptotical $\mathrm{mod}\, O(h^2)$ on $L_2(R^n)$. This representation is characterized by the following properties:

(1) If $g(\xi, p_0)\to(x, p)$ has no focal points, i.e., its Lagrangian graph is diffeomorphically projected on the plane (ξ, x), then $K(g)$ is an integral operator with the kernel

$$(-2\pi ih)^{-n/2}\left|\frac{\partial^2 S}{\partial\xi\,\partial x}\right|^{1/2} e^{-iS(x,\xi)/h},$$

where $S(x, \xi)$ is the producing function,

(2) $K(g_1\otimes g_2) = K(g_1)\otimes K(g_2)$,

(3) $f_2\left(x, \dfrac{h}{i}D\right)K(g)f_1\left(x, \dfrac{h}{i}D\right) = 0$, if $g(\mathrm{supp} f_1)\cap(\mathrm{supp} f_2) = \emptyset$.

Denote $H_I(p) = \sum\limits_{i\in I}\dfrac{p_i^2}{2}$, $I\subset\{1, \ldots, n\}$. If g is such a canonical transformation that for some $t>0$ and $I\subset\{1, \ldots, n\}$ the mapping $g_{H_I}^{-\tau}g$ has no focal points for $\tau\leqslant t$, then properties (1), (2) imply that $K(g)$ has the kernel

$$\mathscr{K}^g(x, \xi) = \tau^{-k/2}(-2\pi ih)^{-(n+k)/2}\int\limits_{R^k}\left|\frac{\partial^2 S(\xi, y)}{\partial\xi\,\partial y}\right|^{1/2}\times$$
$$\times\exp\left\{-\frac{i}{h}\left(S(\xi, y)+\frac{1}{2\tau}\sum_{j\in I}(x_j-\eta_j)^2\right\}d\eta, \qquad (1.7)$$

---

[1] In the same way the canonical operator is introduced in the general problem of asymptotics with respect to smoothness (Maslov, 1965, [103]) and for the parametrix (Maslov, 1967, [105]; Duistermaat and Hörmander, 1972, [38]; Hörmander, 1971, [56]). In the last case the canonical operator is often called the integral Fourier operator. Moreover, the canonical operator is defined on a Lagrangian manifold [103], which enables us to solve stationary problems.

where $y_j = \eta_j$ for $j \in I$, and $y_j = x_j$ for $j \notin I$, and $S(\xi, y)$ is the producing function of the mapping $g_{H_I}^{-\tau} g$. Property (3) implies, that $K(g)$ can be represented as a zero-dimensional operator cocycle on the Lagrangian graph Λ_g of the mapping g: to any open set $\Omega \subset \Lambda_g$ corresponds a local kernel $\mathscr{K}_\Omega(x, \xi)$, $\mathscr{K}_{\Lambda g} = \mathscr{K}^g(x, \xi)$. This fact and the statement given below enable us to give the local version of formula (1.7).

LEMMA. *Any point $a \in \Lambda_g$ is non-singular with respect to the projection on a coordinate plane $\{p_0 = 0; x_i = 0, i \in I; p_i = 0, i \notin I\}$. If the point $a \in \Lambda_g$ is non-singular with respect to π_I, then for sufficiently small $\tau > 0$ the point $g_{H_I}^{-\tau} a \in \Lambda_{g_{H_I}^{-\tau} g}$ is non-singular with respect to the projection π on the plane $\{p_0 = 0, p = 0\}$.*

Let the point $a \in \Lambda_g$ be non-singular with respect to π_I. Then for a sufficiently small neighbourhood Ω of the point a the formula is valid, which is analogous to (1.7):

$$\mathscr{K}_\Omega(x, \xi) = \tau^{-k/2}(-2\pi i h)^{-(n+k)/2} e^{i\gamma} \int\limits_{R^k} \left| \frac{\partial^2 S(\xi, y)}{\partial \xi \, \partial y} \right|^{1/2} \times$$
$$\times \exp\left\{ -\frac{i}{h} \left(S(\xi, y) + \frac{1}{2\tau} \sum_{j \in I} (x_j - \eta_j)^2 \right) \right\} \varphi(y) \, d\eta \tag{1.8}$$

for (x, ξ) close to $\pi(a)$. Here $y_j = \eta_j$ for $j \in I$ and $y_j = x_j$ for $j \notin I$; φ is a "cutting" function equal to unity near $\pi\big(g_{H_I}^{-\tau}(a)\big)$; $S(\xi, y)$ is a local producing function of the mapping $g_{H_I}^{-\tau} g$; $\tau > 0$ is sufficiently small and the real number γ is defined by the Maslov index on the manifold Λ_g.

In the construction of the canonical operator one can use, instead of $g_{H_I}^{-\tau}$, other canonical transformations. In the original definition (Maslov, 1965, [103]) the rotation through an angle of $\pi/2$ with respect to some of the coordinates were used, i.e., $g_{\tilde{H}_I}^{\pi/2}$, where $\tilde{H}_I(x, p) = \frac{1}{2} \sum_{i \in I} (x_i^2 + p_i^2)$. More general canonical transformations were used in [56].

Formula (1.8) presents the method of calculating the asymptotical Green function

$$G(x, \xi, t) \approx \mathscr{K}_{g_H^t}(x, \xi)$$

for equation (1.6). In the neighbourhood of the point a we have

$$G(x, \xi, t) \approx \sum_{a \in \pi^{-1}(x_0, \xi_0)} \mathscr{K}_{\Omega(a)}(x, \xi), \tag{1.9}$$

where $\Omega(a)$ is a small neighbourhood of the point a on $\Lambda_{g_H^t}$.

The asymptotics of the Green function for the tunnel equation (1.6) is given by the same formula as (1.9), where in the expression (1.8) for $\mathscr{K}_\Omega$ the parameter h should be changed into ih and the phase factor $e^{i\gamma}$ should be omitted, the parameter τ should be chosen so small as to keep the function $S(\xi, y)$ positive. Then the sum (1.9) is taken only at the points $a = \{\xi_0, L_y[y(0), \dot{y}(0), 0], x_0, L_y[y(t), \dot{y}(t_0), t]\} \in R^{4n}_{\xi,p_0,x,p}$, where L is the Lagrangian corresponding to the Hamiltonian H, and $y(\cdot)$ runs over the set of solutions of the variational problem

$$\int_0^t L(y, \dot{y}, \tau)\, d\tau \to \inf, \qquad y(0) = \xi,\ y(t) = x.$$

The construction described here gives the asymptotics of the Green function uniform in t in the domain $0 < \delta \leqslant t \leqslant T$. The asymptotical solution which is uniform in t on $[0, T]$ has an essentially more complicated form. The proposed method permits minimizing the number of integrals at the focal point.

In the problem of the asymptotics of the Green function the focal points arise only in the case of equations with varying coefficients. As an example consider the Markovian chain

$$u_x^{t+h} = P^+_{x-h} u^t_{x-h} + P^-_{x+h} u^t_{x+h}, \qquad h > 0,$$
$$\{(k, j) \in Z_+ \times Z,\ t = kh,\ x = jh\},$$

where u_x^t is the probability of the chain being at time t in the state x,

$$P_x^\pm = \varphi^\pm|_{x=jh}, \qquad \varphi^+(x) = \frac{1}{4} + \frac{\cos^2 x}{2}, \qquad \varphi^-(x) = \frac{1}{4} + \frac{\sin^2 x}{2}.$$

The following h-pseudodifferential equation corresponds to this Markovian chain:

$$e^{h\partial/\partial t} u = \{e^{-h\partial/\partial x}\varphi^+(x) + e^{h\partial/\partial x}\varphi^-(x)\} u,$$

its Hamiltonian having the form

$$H(x, p) = \ln\{\varphi^+(x) e^p + \varphi^-(x) e^{-p}\}.$$

This Hamiltonian is of the tunnel type. The focal point for g_H^t arises at time $t = \pi\sqrt{3}/2$.

4.3. *Asymptotics of solutions of stationary problems.* An altered construction can be used to obtain the asymptotics of the solution of stationary problems, e.g., of the stationary Schrödinger equation

$$\left(-\frac{h^2}{2}\Delta+V(x)\right)\psi_k = \lambda_k\psi_k, \tag{1.10}$$

where $\psi_k = \psi_k(x) \in L_2(R^n)$ is the eigenfunction, λ_k is the eigenvalue corresponding to ψ_k, and $V(x)$ is a positive smooth function such that $|V(x)| \to\infty$ for $|x|\to\infty$. This function is called potential.

In such problems characteristics of a different type may arise. For example, consider the Schrödinger equation for the quantum oscillator

$$\left(-\frac{h^2}{2}\frac{d^2}{dx^2}+\omega^2\frac{x^2}{2}\right)\psi_k = \lambda_k\psi_k, \qquad \omega = \text{const},\ \omega>0,$$

where k is the number of eigenfunction. To such numbers $k\to\infty$ that $kh\to\mathscr{E}/\omega = \text{const}$ as $h\to 0$ correspond the classical characteristics Λ_k: $\{p^2+\omega^2x^2 = 2\lambda_k\}$, i.e., the closed curves in the phase space. The values of λ_k close to $\mathscr{E}$ are determined by the quantum condition $(2\pi h)^{-1}\oint p\,dq = \lambda_k/(\omega h) = k+1/2$ and the eigenfunctions ψ_k are asymptotically equal to the canonical operator constructed on the curve Λ_k and applied to the function equal to 1.

On the other hand, to the vacuum state (the state for $k = 0$ and $\lambda_k = \omega h/2$) corresponds the eigenfunction $\psi_0 = \exp(-\omega x^2/2h)$, whose logarithmic limit is equal to $S(x) = -\lim\limits_{h\to 0} h\ln\psi_0 = \omega x^2/2$. The function $S(x)$ is the characteristic corresponding to the Hamiltonian $H = p^2/2-\omega^2x^2/2$ with the potential $V(x) = -\omega^2x^2/2$, which is the "inverted" with respect to the classical potential $V(x) = \omega^2x^2/2$. Thus the function ψ_0 can be written in the form

$$\psi_0 = \exp\left(-\frac{1}{h}\int\limits_0^x p\,dx\right),$$

where (p, x) is the solution of the characteristics equation $p^2-\omega^2x^2 = 0$ distinguished by the condition $\int\limits_0^x p\,dx \geqslant 0$.

In the general case the exponential asymptotics of the eigenfunctions corresponding to lower energetic states of the Schrödinger equation (1.10) is defined by the characteristic equation with the Hamiltonian $H(x,p) = p^2/2-V(x)$. This Hamiltonian differs from the known quantum Hamiltonian (1.10) in "turning over" the potential. If the potential $V(x)$ has some points $\xi_0, \xi_1, \ldots, \xi_l$ of the global minimum, then the logarithmic limit of the eigenfunction corresponding to the vacuum state is defined by the solution of the following variational problem:

$$\int (\dot{q}^2/2 - V(q)\,d\tau \to \min_k \inf_t \inf_l,$$

$$l = \{q = q(\tau),\ \tau \in [0, t]|\ q(0) = \xi_k,\ q(t) = x\}.$$

Let the potential $V(x)$ have two points ξ_0, ξ_1 of the global minimum and be a function symmetrical with respect to, for example, the plane $\{x_1 = 0, x_2 = 0, \ldots, x_{n-1} = 0\}$. Then the two lower (vacuum) eigenvalues λ_0 and λ_1 differ by a value exponentially small with respect to the parameter h (whereas the distance between the eigenvalues is a value of the order h). Note that for $\Delta\lambda = \lambda_1 - \lambda_0$ there exists a logarithmic limit

$$-\lim_{h\to 0} h \ln \Delta\lambda = \min_l \int_l p\,dq, \tag{1.11}$$

where $l = \{p = p(t), q = q(t)\}$ is the solution of the Hamilton system

$$\dot{q} = p, \quad \dot{p} = V_x(q)$$

satisfying the conditions

$$q|_{t=-\infty} = \xi_0, \quad q|_{t=\infty} = \xi_1.$$

The trajectories l minimizing the integral in (1.11) correspond to the so-called instantons. The role played by instantons in the splitting of energetic levels of vacuum states is intensively studied in physical literature at present ([171], [164], [11]).

Note in conclusion that the tunnel canonical operator can be applied as well in nonlinear problems to the Boltzmann equation in the case of weak collisions, to large deviations, to the system of equations of gas dynamics with small viscosity, to the asymptotics of the shock wave running far ahead, and to some other problems. However, the formation of this asymptotics involves great mathematical difficulties.

5. Characteristics and global asymptotics as $h\to 0$ for equations with nonlocal nonlinearity. Equations of the Vlasov type for the propagation of wave fronts of oscillations. Bicharacteristics defining canonical transformations. Now we consider the fifth generalization of the notion of characteristics. We consider the characteristics of nonlinear h-pseudodifferential equations. Unfortunately, the equations which describe them are usually more complicated than the Hamilton–Jacobi equations. On the other hand, they usually correspond to well-known physical problems, though physically these problems are sometimes not connected with the original equations at all.

We first consider an important class of nonlinear equations of the following form. Assume that the symbol of a linear h-pseudodifferential operator depends on the solution of the given equation in some special way. Namely, it depends only on the square of its module, to which some linear smoothing operator has already been applied.

In this case the original symbol for a wide class of solutions has a limit as h tends to zero.

Owing to this fact it is possible to solve the preliminary problem of the propagation of the wave front of oscillations. This problem in the linear situation is similar to the problem of the Hörmander wave front for the singular solution. Then the global asymptotical solution can be constructed with rather general initial data. As an example we consider an analogue of the bicharacteristics system for the nonlinear Schrödinger equation with non-local interaction:

$$ih\frac{\partial\psi}{\partial t} = -\frac{h^2}{2}\Delta\psi - \psi\int K(x-\xi)|\psi(\xi, t)|^2 d\xi, \tag{1.12}$$

where $x = (x_1, \ldots, x_n)$, $\xi = (\xi_1, \ldots, \xi_n)$, $h \to 0$. Here the bicharacteristics are defined by the systems of partial integro-differential equations

$$\begin{gathered}
\frac{\partial Q(x,p,t)}{\partial t} = \Xi(x,p,t), \quad Q(x,p,0) = x, \\
\frac{\partial \Xi(x,p,t)}{\partial t} = \iint K'\big(Q(x,p,t) - Q(\xi,\eta,t)\big) f_0(\xi,\eta)\, d\xi\, d\eta, \\
\Xi(x,p,0) = p, \quad p = (p_1, \ldots, p_n), \quad Q = (Q_1, \ldots, Q_n), \\
\Xi = (\Xi_1, \ldots, \Xi_n), \quad K'(z) = \left(\frac{\partial K}{\partial z_1}, \ldots, \frac{\partial K}{\partial z_n}\right),
\end{gathered} \tag{1.13}$$

where f_0 is equal to the weak limit, as $h \to 0$, of the expression

$$\left(\frac{1}{2\pi h}\right)^n \int e^{ip(x-\xi)/h}\psi(\xi, 0)\bar{\psi}(x, 0)\, d\xi,$$

i.e., of the density of oscillation of the function $\psi|_{t=0}$.

It turns out that *the solution of problem* (1.13) *defines the canonical transformation* $g(t)$: $(x, p) \to (Q, \Xi)$. In this case the function $f_0(g(t)^{-1}(x, p))$ is a solution of the classical Vlasov equation for a self-consistent field ([111]). The solution asymptotics for the initial equation (1.12) can be calculated by the formula

$$\psi = K(g(t))\psi|_{t=0},$$

where K is the canonical operator defined on the Lagrangian graph of the canonical transformation $g(t)$ (Maslov, 1976, [111], [115]). The asymptotics of the solutions of spectral problems for equations with integral nonlinearities are also calculated by means of the canonical operator, in particular the asymptotics of soliton-like solutions for the Ukawa and Choqward equations (Karasev and Maslov, 1979, [64], [65]; Chernykh, 1982, [21]). In the case of the Ukawa equation the calculation of the characteristics is reduced to the classical Blodchett–Langmur equation for the spherical diode ([89]). In the case of the one-dimensional Choquard equation the characteristics are defined by the well-known Dowson ([25]) solution of the equation for the self-consistent field.

Such equations with non-local nonlinearity show in the most complicated physical situation the advantages of the general operator approach to the characteristics. The second part of the book "Algebras with general commutation relations and their applications" by Karasev, Maslov, and Nazailinskii, published in 1979, is entirely devoted to the operator approach to equations of this type.

We have followed all the basic paths of generalization of the notion of characteristics. The general definition contains complex characteristics, which in the case of pure oscillations or in the case of pure damping turn to be real. In general case one can consider an operator-valued n-dimensional symbol and the n-tuple of non-commuting operators, which should be substituted into this symbol. Besides, the symbol may depend, generally speaking, on the solution itself, which results in a nonlinear equation.

Note that the characteristics contain more information than the asymptotical behaviour of the solution of this problem. Just as the classical characteristics, they help to classify the problems and to specify the problem itself, which is absolutely necessary when we consider a new area in mathematical physics. For example, they help a great deal in problems of microelectronics, optical electronics and other problems arising in the electronic computer construction.

Part II. Characteristics of nonlinear equations

1. Linear equations with rapidly oscillating coefficients. Nonlinear wave equations whose characteristics are equations of the dynamics of gases. Characteristics equations for nonlinear wave superposition are equations of the dynamics of a gas mixture. It is well known that the consideration of linear problems with unsmooth coefficients may sometimes give a start to proving the existence theorems for nonlinear equations. For example, penetrating investigations made by Bony, 1981, [15] into problems of the parametrix construction for pseudodifferential equations with an unsmooth symbol led him to the existence theorems for nonlinear equations. The parametrix for the problem with unsmooth coefficients is closely connected with the problem of asymptotics for pseudodifferential equations with rapidly oscillating symbols. The solution of the last problem is evidently an important phase in the investigation of oscillating solutions of the semi-linear equations. There is no doubt that delicate methods used by Bakhvalov, 1974, [9], De Giorgi and Spagnolo, 1973, [26], Lions, 1980, [98], Oleinik and others ([15], [84], [104], [79]) in problems of equations with rapidly oscillating coefficients will be used in problems of semi-linear equations with a parameter. Here we consider in the most general way the analogy between these two groups of problems.

First we deal with the solution asymptotics of the equations with rapidly oscillating coefficients. For example, consider the Schrödinger equation with the rapidly oscillating potential

$$ih\frac{\partial\psi}{\partial t} = -\frac{h^2}{2}\frac{\partial^2\psi}{\partial x^2} + V\left(\frac{\Phi_1(x)}{h}, \ldots, \frac{\Phi_l(x)}{h}, x\right)\psi, \tag{2.1}$$

where $\Phi_j \in C^\infty$, $V(\tau_1, \ldots, \tau_l, x) \in C^\infty$ and V is 2π-periodic with respect to the variables $\tau = (\tau_1, \ldots, \tau_l)$. The solution of (2.1) can be represented in the form

$$\psi = W\left(\frac{\Phi_1(x)}{h}, \ldots, \frac{\Phi_l(x)}{h}, x, t, h\right), \tag{2.2}$$

where $W(\tau_i, \ldots, \tau_l, x, t, h)$ is a 2π-periodical function with respect to its first arguments which, generaly speaking, depends irregularly on h and satisfies the equation:

$$ih\frac{\partial\psi}{\partial t} = \left(\frac{h}{i}D_x - i\sum_{j=1}^{l}\frac{\partial\Phi_j(x)}{\partial x}\frac{\partial}{\partial\tau_j}\right)^2 W + V(\tau_1, \ldots, \tau_l, x)W. \tag{2.3}$$

It is an h-differential equation with the operator-valued symbol, since the small parameter appears by the derivatives with respect to part of the arguments (such equations were considered in Part II.2). The symbol of this equation is the operator

$$L(x, p) = \left(p - i \sum_{j=1}^{l} \frac{\partial \Phi_j(x)}{\partial x} \frac{\partial}{\partial \tau_j}\right)^2 + V(\tau_i, \ldots, \tau_l, x),$$

which is defined on the torus T^l for fixed x, p.

For example, let $H(x, p)$ be the one-tuple eigenvalue of this operator and let $\chi(x, p, \tau_1, \ldots, \tau_l)$ be the eigenfunction corresponding to it. Then the expression

$$\left[K(g_H^t) \circ \chi\left(x, \frac{h}{i} D, \tau_1, \ldots, \tau_l\right) \psi_0(x)\right]\Big|_{\tau_j = \Phi_j(x)/h},$$

where $\psi_0 \in C_0^\infty$ is the solution asymptotics for the initial equation (2.1).

The problem of eigenvalues and eigenfunctions of the operator $L(x, p)$ was essentially simplified by Novikov, 1974, [133], and others ([59], [36]) in the case where the gradients of the phase $\Phi_1, \Phi_2, \ldots, \Phi_l$ are colinear and the function $\omega = V\left(\eta\left|\frac{\partial \Phi_1}{\partial x}\right|, \eta\left|\frac{\partial \Phi_2}{\partial x}\right|, \ldots, \eta\left|\frac{\partial \Phi_l}{\partial x}\right|, x\right)$ is the finite gap potential of the operator $-d^2/d\eta^2 + \omega$ for every x. Then the eigenfunctions and the eigenvalues of the operator $L(x, p)$ can be expressed in terms of the Riemann θ-function, and the solution asymptotics for the initial equation (2.1) has a beautiful geometrical interpretation.

Nonlinear equations with small dispersion also have rapidly oscillating solutions analogous to (2.2) (Whitham, 1965, [165], [166]). Such equations include, in particular, the nonlinear wave equation $h^2(\partial^2 u/\partial t^2 - \Delta u) + g(u, x) = 0$, the Korteweg–de Vries equation and others. The difference from the linear case is that the functions Φ_j (phases) in (2.2) are not given in advance, as can be shown on the example of the following equation:

$$h^2\left(\frac{\partial^2 u}{\partial t^2} - \Delta u\right) + a(x)\operatorname{sh} u = 0, \quad x \in R^3, \quad a \in C^\infty. \tag{2.4}$$

The function

$$u = f\left(\frac{S(x, t)}{h}, x, t, h\right) \tag{2.5}$$

($S \in C^\infty, f(\tau, x, t, h) \in C^\infty$ is a 2π-periodic function with respect to the argument τ) is the solution of (2.4) if

$$\left\{\left[\frac{\partial S}{\partial t}\frac{\partial}{\partial \tau}+h\frac{\partial}{\partial t}\right]^2-\left[\frac{\partial S}{\partial x}\frac{\partial}{\partial \tau}+h\frac{\partial}{\partial x}\right]^2\right\}f+a(x)\operatorname{sh} f=0. \tag{2.6}$$

Differing from the linear case, the asymptotical solution f of equation (2.6) is regular with respect to the parameter h and hence, the leading term $f(\tau, x, t, 0)$ satisfies equation (2.6) (ordinary with respect to the argument τ) for $h = 0$. The function $S(x, t)$ and two "constants" of integration (which are functions in variables x, t), on which the function $f(\tau, x, t, 0)$ depends, can be obtained from the 2π-periodicity conditions with respect to the argument τ for the functions $f(\tau, x, t, 0)$ and $\frac{\partial f}{\partial h}(\tau, x, t, 0)$. These conditions lead to the "clutched" system of equations, which is equivalent to a system of relativistic hydromechanics equations without whirls [107]. The asymptotical rapidly oscillating solutions and the corresponding equations of characteristics were studied and applied in Luke, 1966, [99], Zabusky, 1967, [168], Berezin and Karpman, 1966, [14], Miura and Kruskal, 1974, [129], Povsner, 1974, [138], Scott 1970, [144], Ostrovskii, 1966, [135], Pelinovskii, 1982, [136], Gurevich and Pitaevskii, 1973, [52] and others.

In the wave equations connected with the interference and interaction of waves which arise, in particular, by reflection from the boundary there exist multi-phase solutions which are nonlinear superpositions of solutions (2.5). They have the form

$$u = \mathscr{F}\left(\frac{S_1(x, t)}{h}, \ldots, \frac{S_l(x, t)}{h}, x, t, h\right),$$

where $\mathscr{F}(\tau, x, t, h)$ is a smooth function with respect to the arguments $\tau = (\tau_1, \ldots, \tau_l), x, t,$ being 2π-periodic with respect to each τ_j and, generally speaking, irregularly depending on h. In this case the leading term $\mathscr{F}(\tau, x, t, 0)$ satisfies the equation with partial derivatives with respect to the variables $\tau_1, \ldots, \tau_l$ [1]. An effective solution of this equation became possible after the papers by Novikov, 1974, [133] and other authors [133], [59], [36], [78], [82], [90], [102], [125], [155], [131], which were devoted to the finite gap almost periodic solutions. The connection between the finite gap almost periodic solutions and the function $\mathscr{F}(\tau, x, t, 0)$ was investigated by means of the problem of the reflection from the boundary for the nonlinear equation (2.4) (Dobrokhotov and Maslov, 1979, [30], [31]). It turned out that in this problem the function $\mathscr{F}(\tau, x, t, 0)$ is expressed in terms of the almost periodic finite gap solution of the sine-Gordon equation $\partial^2 v/\partial\eta^2 - \partial^2 v/\partial\xi^2 + \sin v = 0$.

Thus the finite gap solutions define the "superpositions" of the asymptotical solutions of the multidimensional equations (2.4) with variable coefficients and a small parameter. The same correspondence between the finite gap almost periodic solutions and multi-phase asymptotics was used for constructing the solutions of the Korteweg-de Vries equation (Dobrokhotov and Maslov, 1980, [32], [34], Flaschka, Forest and McLaughlin, 1980, [40]).

Flaschka, Forest and McLaughlin have reduced the corresponding system of characteristics to an elegant geometrically invariant form, and Novikov and Dubrovin have recently shown that it is equivalent to the equation of the dynamics of a gas mixture, and have constructed for this system a beautiful Hamiltonian formalism.

The corrections of the leading term of the asymptotical solution of the corresponding nonlinear equation (the higher terms of the asymptotical expansion) are defined from the linear equation (with different right sides). Its coefficients are functions in the leading term of the asymptotical solution and oscillate rapidly. By applying the procedure given at the beginning of this section, this equation is reduced to an equation with an operator-valued symbol. Its eigenvalues for $l \geqslant 2$, as a rule, change multiplicity, and for the construction of the solutions [34] the methods developed by Kucherenko, 1974, [84], in problems of non-strictly hyperbolic systems of equations are applied.

2. Nonlinear equations with complex characteristics. Wave superposition law and its analogy with the soliton addition law. In the case where the characteristics of nonlinear equations are complex, the nonlinear WKB method (the Whitham method) together with the theorem on complex germ leads to essentially simpler answers compared with the case of real characteristics. The complex characteristics arise, in particular, in the problem with the impedance boundary condition for the class of elliptic nonlinear equations

$$h^2 \sum_{j=1}^{n} \frac{\partial}{\partial x_j} a_{ij} \frac{\partial}{\partial x_i} u = g(u, x), \qquad x \in \Omega \subset R^n, \tag{2.7}$$

where $a_{ij}(x) \in C^\infty$, $\|a_{ij}\| > 0$, $g(u, x)$ is a smooth function in the arguments x and entire in the argument u, $g(0, x) = 0$, $\partial g/\partial u(0, x) > 0$ (such equations are known, in particular, in the theory of semiconductors [72], [3]). In this case the characteristics coincide with the characteristics of the corresponding linear equation. For example, in problem (2.7) the characteristics are the same for all the functions having the same part $g'_u(0, x)u$ linear in u. Highly interesting and general is the "law of nonlinear superposition" of the solution asymptotics, which can be formulated by using the results [93], [94] of the Dirichlet series theory. For example, for equation (2.5) this law does not depend on the number of variables x or on the nonlinear component with respect to the argument u of the function $g(u, x)$ (Dobrokhotov and Maslov, 1980, [33]). The application of this law enables us, by using the "one-phase" always complex solutions $f(S(x)/h, x)$, $\operatorname{Im} S > 0$, defined by one-dimensional Dirichlet series for the function $f(\tau, x)$, to reconstruct the "multiphase" already both complex and real-valued solutions $F\big(S_1(x)/h, S_2(x)/h, \ldots, S_l(x)/h\big)$. They are defined by means of the multidimensional Dirichlet series for the function $F(\tau_1, \ldots, \tau_l, x)$. In the case of $g = \operatorname{sh} u$ the Dirichlet series are summed up, the function $F(\tau_1, \ldots, \tau_l, x)$ can be expressed in terms of the multi-soliton solutions of the sine-Gordon equation, and the "superposition" formulas for the Dirichlet series turn out to be equivalent to the "superposition" formulas for solitons (Dobrokhotov and Maslov, 1977, [113], [29]).

Though the complex characteristics of equation (2.5) coincide with the characteristics of the corresponding linear equations, the nonlinear effects in the solution asymptotics of these equations are essential. For the equation of a semiconductor, for example, this results in a considerable increase of conductivity in the neighbourhood of some curve on its surface [28].

In conclusion, we give the solution asymptotics for a bundary problem with the impedance condition for the equation of an electrically neutral semiconductor $h^2 \Delta u = \operatorname{sh} u$ in a half-infinite straight elliptic cylinder

$$\Omega = \left\{ x \in R^3 \,\middle|\, \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \leqslant 1,\ x_3 \geqslant 0 \right\}, \qquad a < b\sqrt{2}.$$

They have the form [28]

$$u_n = 2 \ln \frac{1 + \varphi_1 e^{is_l^k/h} + \varphi_2 e^{is_2^k/h} + \frac{1}{4}\varphi_1\varphi_2(\lambda_k^2 - 1)\lambda_k^{-2} e^{i(s_1^k - s_2^k)/h}}{1 - \varphi_1 e^{is_1^k/h} + \varphi_2 e^{is_2^k/h} + \frac{1}{4}\varphi_1\varphi_2(\lambda_k^2 - 1)\lambda_k^{-2} e^{i(s_1^k - s_2^k)/h}},$$

where the characteristics have the form

$$s_1^k = -\bar{s}_2^k = \sqrt{\lambda_k^2-1}\left(x_1+a+\frac{\beta x_1^2}{2\left((1+\beta(x_1+1)\right)}\right)+i\lambda_k x_3, \qquad k\sim\frac{1}{h},\ k\in Z_+,$$

$$\varphi_1=\bar{\varphi}_2=\sqrt{1+\beta(x_1+a)}, \qquad \beta=(-a+i\sqrt{b^2-a^2})/b^2$$

and

$$\lambda_k=\frac{1}{h}\sqrt{1+\frac{h^2}{4a^2}(\pi k+\tfrac{1}{2}\arccos\left(1-\frac{2a^2}{b^2}\right))}$$

is the impedance.

If in the formula for u_k we set

$$s_1^k=-i(\sqrt{1+V^2}\xi+V\eta), \qquad s_2^k=-i(\sqrt{1+V^2}\xi-V\eta),$$

$$\varphi_1=\varphi_2=\frac{iV}{(1+V^2)^{1/4}}, \qquad \lambda_k=iV,\ V=\text{const},$$

then the function u_k modulo the factor i coincides with the soliton–antisoliton solution of the sine-Gordon equation.

3. Singularity propagation in nonlinear equations. Conditions of the Hugoniot type are the analogue of characteristics equations for this problem. Singularities of branching type. Solutions with finite outliers at single points.

3.1. *Singularities in nonlinear equations.* In linear hyperbolic equations the characteristics define the dynamics of the discontinuous part of the solution. In modern educational literature this fact is described by means of the theory of distributions and the so-called ray expansions. The solution substituted in the equation is of the form:

$$\varphi_0+\theta(s)\varphi_1+s\theta(s)\varphi_2+\dots,$$

where $\theta(\tau)$ is the Heaviside function, $s=s(x,t)\in C^\infty$, $\nabla_x s|_{s=0}\neq 0$, $\varphi_i\in C_0^\infty$, the surface of zeros of the function defines the discontinuity surface. The equations for the functions s and for the coefficients $\varphi_i=\varphi_i(x,t)$ of this series are obtained by equating successively to zero the coefficients at the distributions $\delta(\tau)$, $\theta(\tau)$, $\tau\theta(\tau)$, etc. This procedure is transferred to the semi-linear hyperbolic equations, written in the so-called divergence form. The possibility of such generalization is based on the fact that such a series forms a commutative algebra with respect to usual multiplication. Consequently, the powers of this series are again a series of the same type, and the derivative of this series will again be a series of the same type (with the addition of δ-function only).

In the nonlinear case, however, the equations for the functions s and φ_i do not split, but are pairwise recurrently clutched (with the same "clutching coefficient" [116]). The first two of these equations are called conditions of the Hugoniot type and are the generalization of the notion of characteristics equation for the semi-linear hyperbolic equation.

However, such a generalization for the nonlinear case does not suit arbitrary types of solution singularities, since the following two conditions should be satisfied for the corresponding series. Firstly, these series should form a commutative algebra with respect to multiplication, secondly, the smoothness of the subsequent terms should increase by unity (if the initial equation is a differential one).

It turns out that only two algebras satisfy these conditions in the situation of "general position". If we admit the Ivanov concept, [60], [61], about the multiplication of distributions, then only three algebras are possible [122].

The first algebra corresponds to the solution describing shock waves, the second corresponds to detonation waves, and the third describes infinitely narrow solitons (which will be discussed below).

Nevertheless, when the nonlinearity in the equation has the character of a power, then for special coefficients and solutions vanishing on one side of the discontinuity surface ($\varphi_0 \equiv 0$) algebras of the type $\theta(s)s^a\sum_{k\geqslant 0} s^k\varphi_k$, where a depends on the nonlinearity degree, are possible.

The propagation of singularities of this type is possible not only for hyperbolic equations. In this case the leading terms of the series after the substitution in the equation are defined not only by the derivatives but also by the powers of the unknown function. Then the equation of the characteristics is of the Hugoniot type, and in this case it is nonstandard. Such propagation of singularities for nonlinear parabolic equations has a physical sense and was considered in [169], [10].

Note in conclusion that the same semi-linear differential equation can be reduced to different divergence forms by multiplication by different powers of the dependent variable, which leads to different conditions of the Hugoniot type for the same semi-linear equation. However, it seems that there is a contradiction. Singular solutions arise as limits for more exact solutions with a small parameter ε by the linear differential operator [57], [22], [58], [114]. Therefore to whatever distribution the solution of this more exact equation converges as $\varepsilon\to 0$, the summands containing ε tend in the equation to zero as distributions (in the weak sense). This property does not allow us to multiply the equation by the powers of the dependent variable and to pass to another divergence form.

Thus, from the point of view of singular solutions it is necessary to write the semi-linear equation in the right divergence form. Then the solution of the limit equation does not depend on the initial equation. This fact is also valid for infinitely narrow solitons.

Nevertheless, the generalization of the characteristics responsible for the propagation of the solution singularities of semilinear equations depends in general on the type of the equations with a small parameter from which these semi-linear equations were obtained by passing to the limit. Such limits were intensively studied in [57], [22], [58], [114]. They are closely connected with the conditions of the Hugoniot type.

3.2. *Infinitely narrow solitons.* Infinitely narrow solitons are rather unusual discontinuity solutions. They arise as limits of some solutions, for example of the Korteweg-de Vries equation with a small dispersion h

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+h^2\frac{\partial^3 u}{\partial x^3}=0,$$

where $h\to 0$. This equation has soliton solutions of the form

$$u=B+A\cosh^{-2}\left(\frac{C(x-Vt)}{h}\right), \tag{2.8}$$

where A, B, C, V are constants. As $h\to 0$ such a solution tends to the function which is equal everywhere to B, but at one point $x=Vt$ this function is equal to $B+A$. We denote this function by $A\delta_1(x-Vt)+B$ and call it the infinitely narrow soliton. Such solutions were not considered in the linear theory. We consider for the Korteweg-de Vries equation the Cauchy problem with a one-parameter family of initial data (2.8), where $t=\tau$ and $B=\Phi_0(x)\in C_0^\infty$. The solution of this problem has the form $\Phi(x,t)+A(x,t)\delta_1\big(s(x,t)-\tau\big)$ as $h\to 0$. Here $\Phi(x,t)$ satisfies the limit equation, and $s(x,t)$ and $A(x,t)$ are connected by

$$\begin{gathered}3\frac{\partial s}{\partial t}+(A+3\Phi)\frac{\partial s}{\partial x}=0, \qquad s|_{t=0}=x/V,\\ \frac{\partial}{\partial t}\big(A(x,t)+2\Phi(x,t)\big)=0.\end{gathered} \tag{2.9}$$

These conditions are analogous to those of the Hugoniot type. It is natural to consider them as the equations of characteristics for the limit equation, corresponding to the propagation of an "infinitely narrow soliton" δ_1 on a variable smooth background. The background $\Phi(x, t)$ in the rapidly oscillating case was calculated by Lax, 1979, [91]. Note that these conditions are applicable to the more general multi-soliton case [119], [118], [123]. Like conditions of the Hugoniot type, they give a good qualitative and quantitative description of the initial equation also for not very small h. Numerical experiments confirm this fact [119]. The second expression in (2.9) essentially depends on the form of the initial equation with a small parameter. Thus the second condition should be obtained from the asymptotics. On the other hand, the consideration of solution asymptotics from this point of view enables us to simplify essentially, and hence to substantiate ([119], [118], [123], [179]) the asymptotical formulas obtained, for example, in physical papers [63], [75], [45], [136], [88], [69], [159], [132], [126]. In particular, even condition (2.9) has been unknown in physical literature.

This approach can be applied also to h-pseudodifferential semi-linear equations (in particular, to the Toda lattice equation, to difference equations, etc.). Considerations from the general point of view enable us to transfer the results obtained for some equations to others. For example, the important results obtained by Rusanov and Bezmenov, 1980 [119], for difference schemes have been used in the problem on asymptotics for nonlinear oscillations of a lattice (Maslov and Mosolov, 1983).

We have discussed the propagation of the singularities of solutions of quasi-linear equations, singularities whose support is of codimension 1. For the propagation of the singularities with the support at a point one can also write conditions of the Hugoniot type, namely, the characteristics equations. They well describe qualitatively, for example, some physical phenomena which arise near the centre of the typhoon (the so-called eye of the typhoon) if we consider the typhoon as the propagation of the singularities for some complicated system of quasi-linear equations.

References

[1] Ablovitz M. A. and Benny D. J., The Evolution of Multiphase Models for Nonlinear Dispersive Waves, *Stud. Appl. Math.* **49** (3) (1970), pp. 225–238.

[2] Analyse Lagrangienne et mecanique quantique, *Seminaire du College de France, 1976–1977 of R.C.P.* **25**, I.R.M.A., Strasbourg, 1978.

[3] Anselm A. N., Introduction to the Theory of semiconductors, Nauka, Moscow, 1978 (in Russian).
[4] Arnold V. I., Integrals of Rapidly Oscillating Functions and Singularities of Lagrangian Manifold Projections, *Funkts. Anal.* **6** (3) (1972) (in Russian).
[5] Babich M. V., Asymptotics with Respect to a Large Parameter for the Fock–Klein–Gordon Equation with a Discontinuous Initial Condition, *Zap. nauch. sem. LOMI* **115** (1982), pp. 16–22 (in Russian).
[6] Babich V. M., and Buldyrev V. S., Asymptotic Methods in the Problems of Short-Wave Diffraction, Nauka, Moscow, 1972 (in Russian).
[7] Bagrov V. V. and Belov V. V., Semi-Classial Trajectory-Coherent States of a Relativistic Spliness Particle in an Arbitrary Electromagnetic Field, *Izv. VUZov, Physics* **46** (1982) (in Russian).
[8] Bagrov V. G., Belov V. V., and Ternov I. M., Semi-Classical Trajectory-Coherent States of a Non-Relativistic Electron in an Arbitrary Electromagnetic Field, *TME* **50** (3) (1982) (in Russian).
[9] Bakhvalov N. S., Averaged Characteristics of Bodies with a Periodic Structure, *DAN SSSR* **218** (5) (1974), pp. 1046–1048 (in Russian).
[10] Barenblatt G. I. and Monin A. S., On the Possible Mechanics of the Appearance of Discoid Formings in Atmosphere, *DAN SSSR* **246** (4) (1979), p. 834 (in Russian)
[11] Belavin A. A., Polyakov A. M., Schwartz A. S., and Tyupkin Yu. S., Pseudo Particle Solutions of Yang–Mills-Equations, *Phys. Lett. Ser. B* **59** (1975), p. 85.
[12] Bensoussan A., Lions J.-L., and Papa Nicolaou G., *Asymptotic Analysis for Periodic Structures*, North-Holl. Publ. Comp., 1978.
[13] Berezin F. A., Quantization in Complex Symplectic Spaces, *Izv. AN SSSR, ser. matem.* **39** (2) (1975), pp. 363–402 (in Russian).
[14] Berezin Yu. A. and Karpman V. I., On Nonlinear Evolution of Perturbations in Plasma and Other Dispersing Media, *Zhurnal exper. i teor. fys.* **51** (3) (1966), pp. 1557–1568 (in Russian).
[15] Bony J.-M., Calcul symbolique et propagation des singularités pour les equations partielles non lineaires, *Ann. Sci. Ec. Norm. Super.* **14** (2) (1981), pp. 209–246.
[16] Borel A. and Hirzebruch F., Characteristic Classes and Homogeneous Spaces, II, *Amer. J. Math.* **81** (2) (1959), pp. 315–382.
[17] Borovkov A. A., Boundary Conditions for the Problem for Random Wandering and Large Deviations in Functional Spaces, *Teorya veroyatn. i primen.* **12** (4) (1967), pp. 635–654 (in Russian).
[18] Boutet de Monvel L., and Guillemin V., The Spectral Theory of Toeplitz Operators, *Ann. Math. Stud.* **99** (1981), 161 pp.
[19] Buslayev V. S., A Generating Integral and Maslov's Canonical Operator in the WKB Method, *Funk. Analiz i Prilozh.* **3** (3) (1969), pp. 17–31 (in Russian).
[20] Chebotarev A. M. and Maslov V. P., 1982, On the Second Term of the Logarithm Asymptotics of the Path Integrals, *Itogi Nauki i Tekhniki. Teoria veroyatn., matem. statist., teor. kybernetiki* **19**, VINITI, Moscow, pp. 127–154.
[21] Cernykh S. I., The Semi-Classical Particle in a One-Dimensional Self-Coherent Field, *TMF* **52** (3) (1982), pp. 491–494 (in Russian).
[22] Courant R., *Partial Differential Equations*, N. Y.–London, 1962.
[23] Czyż J., On Geometric Quantization and Its Connections with the Maslov Theory, *Reports Math. Phys.* **15** (1) (1979), pp. 57–97.

[24] Danilov V. G. and Maslov V. P., Quasi-Invertibility of Functions in Ordered Operators in the Theory of Pseudodifferential Equations, *Sovremen. Probl. Mat.* (*Itogi nauki i tekhniki*) **6**, VINITI, Moscow, 1975, pp. 5–130 (in Russian).

[25] Dawson J. M., Nonlinear Electron Oscillations in a Cold Plasma, *Phys. Rev.* **113** (2) (1959), pp. 383–386.

[26] De Giorgi E. and Spagnolo S., Sulla convergenzi degli integrali dell' energia per operatori eliddici del 2 ordime, *Boll. Un. Math. Ital.* **4** (8) (1973), pp. 391–411.

[27] Dobrokhotov S. Yu. and Maslov V. P., Some Applications of the Complex Germ Theory to Equations with a Small Parameter, *Sovremen. Probl. Mat.* (*Itogi nauki i tekhniki*) **5** VINITI, Moscow, 1975, pp. 141–211 (in Russian).

[28] Dobrokhotov S. Yu. and Maslov V. P., Multidimensional Effects in the Steklov Type Boundary Problem For a Nonlinear Equation of a Semiconductor, *Uspekhi Mat. Nauk* **33** (5) (1978), p. 142 (in Russian).

[29] Dobrokhotov S. Yu. and Maslov V. P., Spectral Boundary Problem Asymptotics for a Nonlinear Equation of a Semiconductor, *DAN SSSR* **243** (4) (1978), pp. 987–900 (in Russian).

[30] Dobrokhotov S. Yu. and Maslov V. P., Solution Asymptotics of a Mixed Problem for the Nonlinear Equation $h^2\Box u + a \operatorname{sh} u = 0$, *Uspekhi Mat. Nauk* **34** (3) (1979), pp. 225–226 (in Russian).

[31] Dobrokhotov S. Yu. and Maslov V. P., Boundary Reflection Problem for the Equation $h^2\Box u + a \operatorname{sh} u = 0$ and finite Zone Almost Periodic Solutions, *Funk. Analiz i Prilozh.* **13** (3) (1979), pp. 79–80 (in Russian).

[32] Dobrokhotov S. Yu. and Maslov V. P., Finite Zone Almost Periodic Solutions in WKB-Approximations, *Sovremen. Probl. Mat.* (*Itogi nauki i tekhniki*) **15**, VINITI, Moscow, 1980, pp. 3–94.

[33] Dobrokhotov S. Yu. and Maslov V. P., Multidimensional Nonlinear Spectral Boundary Problems and Supersposition of Their Asymptotical Solutions According to the Soliton Summing Law. In: *Partial Differential Equations*, Nauka, Novosibirsk, 1980, pp. 1656–1674 (in Russian).

[34] Dobrokhotov S. Yu. and Maslov V. P., Multiphase Asymptotics of Nonlinear Partial Differential Equations with a Small Parameter, *Math. Physics Reviews* **3**, Sov. Sci. Reviews, Over. Publ. Ass. Amsterdam, 1982, pp. 221–311.

[35] Dobrokhotov S. Yu., The Maslov Methods in the Linearized Theory of Gravitation Waves on the Liquid Surface, *DAN SSSR* **269** (1) (1983), pp. 76–80 (in Russian).

[36] Dubrovin B. A., Matveev V. B., and Novikov S. P., Nonlinear Equations of KdV-Type, Finite Zone Linear Operators and Abelian Manifolds, *Uspekhi Mat. Nauk* **31** (1) (1976), pp. 55–136 (in Russian).

[37] Duistermaat J. J., Oscillatory Integrals, Lagrange Immersions and Unfolding of Singularities, *Comm. Pure Math.* **27** (2) (1974), pp. 207–281.

[38] Duistermaat J. J., and Hörmander L., Fourier Integral Operators, II, *Acta Math.* (1972), p. 128.

[39] Fedoriuk M. V., On the Stability in C of the Cauchy Problem for Difference Equations and Partial Differential Equations, *ZhVM i MF* **7** (3) (1967), pp. 510–540 (in Russian).

[40] Flaschka H., Forest M. G., and McLaughlin P. W., Multiphase Averaging and the Inverse Spectral Solution of the Korteweg-de Vries Equation, *Comm. on Pure and Appl. Mat.* **33** (6) (1980), pp. 739–784.

[41] Folland G. B. and Stein E. M., Estimates for the $\bar{\delta}_B$ Complex and Analytic in the Heisenberg Group, *Comm. Pure Appl. Math.* **27** (4) (1974), p. 429.
[42] Gamkrelidze R. V., Chern Cycles of Complex Algebraic Manifolds, *Izv. AN SSSR, ser. Mat.* **20** (1956), pp. 685–706 (in Russian).
[43] Gaveau B., Systems dynamiques associes a certains operateurs hypoelliptiques, *Bull. Sci. Math.* **102** (3) (1978), pp. 203–229.
[44] Grigis A., Propagation des singularités le long de courbes microbicaracteristique pour des operateurs pseudodifferentiels a caractéristiques doubles, I, *Ann. Sci. Ec. Norm. Super.* **15** (1) (1982), p. 147–159.
[45] Grimshaw, Slowly Varying Solitary Waves, *Proc. Roy. Soc. London* **A-368** (1979), pp. 359–388.
[46] Grushin V. V., Hypoelliptic Differential Equations and Pseudodifferential Operators with Operator-Valued Symbols, *Mat. sbornik* **88** (4) (1972), pp. 504–521 (in Russian).
[47] Grushin V. V., Parametrix Construction for Degenerative Elliptic Operators by the Method of Two-Scale Asymptotic Expansions, *Funk. Analiz i Prilozh.* **116** (2) (1977), pp. 76–77 (in Russian).
[48] Guillemin V. and Schäffer D., *Maslov Theory and Singularities*, Cambridge, 1972.
[49] Guillemin V. and Sternberg S., Some Problems in Integral Geometry and Some Related Problems in Microlocal Analysis, *Amer. J. Math.* **101** (1979), pp. 915–955.
[50] Guillemin V. and Uhlmann G., Oscillatory Integrals with Singular Symbols, *Duke Math. J.* **48** (1) (1981), pp. 251–267.
[51] Guillemin V. and Sternberg V., Homogeneous Quantization and Multiplicities of Group Representations, *J. Funct. Anal.* **47** (3) (1982), pp. 344–380.
[52] Gurevich A. V. and Pitaevskij L. P., Decay of Initial Discontinuity in the KdV-Equation, *Letters in ZhETF* **17** (5) (1973), pp. 268–271 (in Russian).
[53] Hagedern G. A., Semi-Classical Quantum Mechanics, III, The Large Order Asymptotics and More General States, *Ann. Phys.* **135** (1) (1981), pp. 58–70.
[54] Helfer B. and Nourrigat J., Hypoellipticite pour des operators sur des groups de Lie nilpotents, *C. R. Acad. Sci.* **AB 287** (6) (1978), pp. A395–A398.
[55] Hess H., On a Geometric Quantization Scheme Generalizing Those of Kostant–Souriau and Czyż, *Lect. Notes Phys.* **139** (1981), pp. 1–35.
[56] Hörmander L., Fourier Integral Operators, I, *Acta Math.* **127** (1971), pp. 79–183.
[57] Hopf E., On the Right Weak Solution of the Cauchy Problem for a Quasilinear Equation of First Order, *J. Mat. Mech.* **19** (6) (1969), pp. 483–487.
[58] Il'in A. M., The Structure of Shock Wave Small Viscosity. In: *Nonlinear Waves*, Nauka, Moscow, 1982, pp. 234–237 (in Russian).
[59] Its A. R. and Matveev V. B., Hill Operators with a Finite Number of Lacunae and Multi-Solution Solutions of KdV-Equations, *Teor. i Mat. Phys.* **23** (1) (1975), pp. 51–67 (in Russian).
[60] Ivanov V. K., Associative Algebra of Simplest Distributions, *Sib. Mat. J.* **20** (4) (1979), pp. 731–740 (in Russian).
[61] Ivanov V. K., On the Multiplication of Homogeneous Distributions in Several Variables, *DAN SSSR* **257** (1) (1981), pp. 29–33 (in Russian).
[62] Ivrii V. Ya., On Semi-Classical Spectral Asymptotics for the Schrödinger Operator on Manifolds with a Boundary and for h-Pseudodifferential Equations, *Funk. Analiz i Prilozh.* **14** (2) (1980), pp. 25–34 (in Russian).

[63] Johnson R. S., On the Development on a Solitary Wave Moving over an Uneven Bottom, *Proc. Camb. Phil. Soc.* **73** (1973), pp. 183–203.

[64] Karasev M. V. and Maslov V. P., Algebras with General Commutation Relations and Their Applications, II, Operator Unitary-Nonlinear Equations, *Sovremen. Probl. Matem.* (*Itogi nauki i tekhniki*) **13**, VINITI, Moscow, 1979, pp. 145–267 (in Russian).

[65] Karasev M. V. and Maslov V. P., Semi-Classical Soliton Solutions of the Hartri Equation. Newtonian Interaction with Screening, *TMF* **40** (2) (1979), pp. 235–244 (in Russian).

[66] Karasev M. V. and Maslov V. P., Global Asymptotic Operators of Regular Representation, *DAN SSSR* **257** (1) (1981), pp. 33–38 (in Russian).

[67] Karasev M. V. and Maslov V. P., Quantization of Symplectic Manifolds with Conic Points, *TMF* **2** (1982), pp. 263–269 (in Russian).

[68] Karasev M. V. and Maslov V. P., Pseudodifferential Operators in General Symplectic Manifolds, *Izv. AN SSSR* **47** (5) (1983), (in Russian).

[69] Karpman V. I. and Maslov E. M., A Perturbation Theory for the Korteweg-de Vries Equation, *Physics Lett.* **60A** (4) (1977), pp. 307–308.

[70] Keller J. B., Diffraction by a Convex Cylinder, *Trans. IRE Ant. and Prop.* **4** (3) (1956).

[71] Khrgian A. H., *Physics of the Atmosphere*, Gidrometeoizdat, Leningrad, 1969 (in Russian).

[72] Kireev P. S., *Physics of Semiconductors*, Vys. Shkola, Moscov, 1975 (in Russian).

[73] Kifer Yu. I., On Asymptotics of Transitional Densities for Processes with Small Diffusion, *Teoriya Veroyatn. i Primenen.* **21** (3) (1976), pp. 572–536 (in Russian).

[74] Kirillov A. A., The Characters of Unitary Representations of Lie Groups, *Funk. Analiz i Prilozh.* **2** (2) (1968), pp. 40–55 (in Russian)

[75] Ko K. and Kuehl H. H., Korteweg-de Vries Solitons in a Slowly Varying Medium, *Phys. Pev. Lett.* **40** (1978), pp. 233–236.

[76] Kogan V. R., Asymptotics of the Laplace–Beltrami Operator on the unit sphere S^n, *Izv. Vuzov, Radiophysics* **12** (11) (1969), (in Russian).

[77] Kostant B., Quantification and Unitary Representations, *Lect. Notes Math.* **170** (1970), pp. 87–208.

[78] Kozel V. A. and Kotlyarov V. P., Almost Periodic Solutions of the Equation $u_{tt}-u_{xx} = \sin u$, *DAN USSR, ser. A* **10** (1976), pp. 878–881 (in Russian).

[79] Kozlov S. M., Averaging of Random Operators, *Mat. sbornik* **109** (2) (1979), pp. 188–202 (in Russian).

[80] Krakhnov A. D., Asymptotics of Eigenvalues for Pseudodifferential Operators and Invariant Tori, *Uspekhi Mat. Nauk* **31** (3) (1976), pp. 217–218 (in Russian).

[81] Kravtsov Yu. A., Complex Rays and Complex Caustics, *Izv. Vuzov, Radiophysics* **10** (1967), p. 1287 (in Russian).

[82] Krichever I. M., Methods of Algebraic Geometry in the Theory of Nonlinear Equations, *Uspekhi Mat. Nauk* **32** (6) (1977), pp. 183–208 (in Russian).

[83] Kucerenko V. V., Maslov Canonical Operator on the Germ of Complex Almost Analytical Manifold, *DAN SSSR* **213** (6) (1973), pp. 1251–1254 (in Russian).

[84] Kucerenko V. V., Asymptotic Solution of the System for $h\to 0$ in the Case of Characteristics with Variable Multiplicity, *Izv. An SSSR, ser. mat.* **38** (3) (1974), pp. 625–662 (in Russian).

[85] Kucerenko V. V., Semi-Classical Assymptotics of the Point Source Function for the Stationary Schrödinger Equation, *TMF* **1** (3) (1969), pp. 384–406 (in Russian).

[86] Kucerenko V. V., Asymptotics of the Cauchy Problem Solution for Equations with Complex Characteristics, *Sovremen. Probl. Mat.* (*Itogi nauki i tekhniki*) **8**, VINITI, Moscow, 1976, pp. 41–135 (in Russian).

[87] Kulikovskii A. G. and Lyubimov G. V., *Magnetic Hydrodynamics*, Nauka, Moscow, 1962.

[88] Kulikovskii A. G. and Reutov V. A., The Motion of Single and Periodic Waves with the Amplitude Close to the Limit in the Layer of a Liquid of Slowly Varying Depth, *Izv. AN SSSR, MZhG* **6** (1976) (in Russian).

[89] Langmuir I. and Blodgett K., Currents Limited by Space Charge between Concentric Spheres, *Phys. Rev.* **24** (1) (1924), pp. 49–59.

[90] Lax P., Periodic Solutions of the Korteweg-de Vries Equation, *Comm. Pure Appl. Math.* **28** (1975), pp. 141–188.

[91] Lax P. D. and Levermore C. D., The Zero Dispersion Limit for the KdV Equation *Proc. Nat. Acad. Sci. USA* **76** (8) (1979), pp. 3602–3606.

[92] Lazutkin V. F., Spectral Degeneration and "Small Denominators" in the Asymptotics of an Eigenfunction of the "Jumping Ball" Type, *Vestnik LGU* **7** (1969), (in Russian).

[93] Leont'ev A. F., *Series of Exponents*, Nauka, Moscow, 1976 (in Russian).

[94] Leont'ev A. F., On the Representation of Analytic Functions by Series of Exponents in the Semi-Cylinder Domain, *Matem. sbornik* **100** (3) (1976), pp. 364–383 (in Russian).

[95] Lieb E. H. and Simon B., The Hartree-Fock Theory for Culomb Systems, *Comm. Math. Phys.* **53** (3) (1977), pp. 185–194.

[96] Leray J., Garding L., and Kotake T., Probléme de Cauchy, *Bull. Soc. Math. France* **92** (1964), pp. 263–361.

[97] Leray J., *Analyse Lagrangienne et mécanique quantique; une structure mathématique apparentée aux developpements asymptotiques et à l'indice de Maslov*, Strasbourg, 1978.

[98] Lions P. L., The Chogward Equation and Related Questions, *Nonlinear Anal.* **4** (6) (1980), pp. 1063–1072.

[99] Like J. C., A Perturbation Method for Nonlinear Dispersive Wave Problems, *Proc. Roy. Soc.* **A292** (1430) (1966), pp. 403–412.

[100] Magneron B., Une généralisation de la notion d'indices de Maslov, *C. R. Acad. Sci.* **AB289** (14) (1979), pp. A683–A686.

[101] Marchenko V. A. and Ostrovskii I. V., Korteweg-de Vries Periodical Problem, *Matem. sbornik* **95** (3) (1974), pp. 331–356 (in Russian).

[102] Maslov V. P., Scattering Problem in Semi-Classical Approximation, *DAN SSSR* **151** (2) (1963), pp. 306–309 (in Russian).

[103] Maslov V. P., *Perturbations Theory and Asymptotical Methods*, MGU, Moscow, 1965 (in Russian).

[104] Maslov V. P., The WKB Method in the Multi-Dimensional case. Appendix in the Russian edition of: Heading J., *An introduction to Phase-Integral Methods*, Mir, Moscow, 1965 (in Russian).

[105] Maslov V. P., On the Regularization of the Cauchy Problem for Pseudodifferential Equations, *DAN SSSR* **177** (6) (1967), pp. 1277–1280 (in Russian).

[106] Maslov V. P., The Theory of Bicharacteristics for Difference Schemes, *Uspekhi Mat. Nauk* **22** (4) (1968), (in Russian).
[107] Maslov V. P., Turning the Heisenberg Equation into the Dynamics Equation for an Ideal Gas and the Quantization of Relativistic Hydrodynamics, *TMF* **3** (1969), pp. 378–383 (in Russian).
[108] Maslov V. P., The Canonical Operator on a Lagrangian Manifold with a Complex Germ and Regularization for Pseudodifferential Operators and Difference Schemes, *DAN SSSR* **195** (3) (1970) (in Russian).
[109] Maslov V. P., *Operational Methods*, Nauka, Moscow, 1973 (in Russian).
[110] Maslov V. P. and Fedoriuk M. B., *Semi-Classical Approximation in Quantum Mechanics*, Nauka, Moscow, 1976 (in Russian).
[111] Maslov V. P., *Complex Markovian Chains and Feynman's Continual Integral*, Nauka, Moscow, 1976 (in Russian).
[112] Maslov V. P., Application of the Ordered Operators Method for Obtaining Exact Solutions, *TMF* **33** (2) (1977), pp. 185–209 (in Russian).
[113] Maslov V. P., *Complex Method WKM in Nonlinear Equations*, Nauka, Moscow, 1977 (in Russian).
[114] Maslov V. P. and Tsupin V. A., Propagation of a Shock Wave in a Proentropic Gas with Small Viscosity, *Sovremen. Probl. Matem.* (*Itogi nauki i tekhniki*) **8**, VINITI, Moscow, 1977, pp. 273–308 (in Russian).
[115] Maslov V. P., The Equations of a Self-Coherent Field, *Sovremen. Probl. Matem.* (*Itogi nauki i tekhniki*) **11**, VINITI, Moscow, 1977 (in Russian).
[116] Maslov V. P., On the Propagation of Shock Waves in an Isoentropic Unviscous Gas, *Sovremen. Probl. Matem.* (*Itogi nauki i tekhniki*) **8**, VINITI, Moscow, 1977, pp. 199–272 (in Russian).
[117] Maslov V. P. and Nazaikinskij V. E., Algebras With General Commutation Relations and Their Applications, I, Pseudodifferential Equations with Growing Coefficients, *Sovremen. Probl. Matem.* (*Itogi nauki i tekhniki*) **13**, VINITI, Moscov, 1979, pp. 5–144 (in Russian).
[118] Maslov V. P. and Tsupin V. A., δ-Formed Solutions, Generalized by Sobolev, of Quasilinear Equations, *Uspekhi Mat. Nauk* **34** (1) (1979), pp. 235–236 (in Russian).
[119] Maslov V. P. and Tsupin V. A., Linear Dependence of Soliton Velocities on Their Base Heights for a KdV-Equation with Small Dispersion, *Uspekhi Mat. Nauk* **35** (5) (1980), p. 218 (in Russian).
[120] Maslov V. P. and Nazaikinskij V. E., Asymptotics for Equations with Singularities in the Characteristics, I, *Izv. AN SSSR* **45** (5) (1981), pp. 1045–1087 (in Russian).
[121] Maslov V. P., Global Exponential Asymptotics for Solutions of Tunnel Type Equations, *DAN SSSR* **261** (5) (1981), pp. 1059–1062 (in Russian).
[122] Maslov V. P., Three Algebras, Corresponding to Non-Smooth Solutions of Systems of Quasilinear Hyperbolic Equations. In: *Distributions and Their Applications in Mathematical Physics, Proc. Inter. Conf.*, Moscow, 1981, pp. 332–333 (in Russian).
[123] Maslov V. P. and Omel'yanov G. A., Asymptotic Solitonlike Solutions of Equations with Small Dispersion, *Uspekhi Mat. Nauk* **36** (3) (1981), pp. 63–126 (in Russian).
[124] Maslov V. P., *Metodes operacianales*, MIR ed., Mascú, 1982.

[125] McKean H. P. and van Moerbeke P., The Spectrum of Hill's Operator, *Invent. Math.* **30** (1975), pp. 217–274.

[126] McLaughlin D. W. and Scott A. C., *Multisolitons Perturbation Theory. Solitons in Action*, Academic Press, New York, 1978.

[127] Melrose R. and Uhlman G., Lagrangian Intersection and the Cauchy Problem, *Comm. Pure Appl. Mat.* **32** (1979), pp. 483–513.

[128] Mishchenko A. S., Sternin B. Yu., and Shatalov V. E., *The Method of Canonical Operator in Applied Mathematics*, MIEM, Moscow, 1974 (in Russian).

[129] Miura R. M. and Kruskal M. D., Application of Nonlinear WKB-Method to Korteweg-de Vries Equation, *SIAM Appl. Math.* **26** (1974), pp. 376–395.

[130] Molchanov S. A., Diffusion Processes and Riemannian Geometry, *Uspekhi Mat. Nauk* **30** (1) (1975), pp. 3–59 (in Russian).

[131] Mumford C. and van Moerbeke P., The Spectrum of Difference Operators and Algebraic Curves, *Acta Math.* (1979).

[132] Newell A. C., Synchronized Solutions, *J. Math. Phys.* **18** (5) (1977), pp. 922–926

[133] Novikov S. P., Periodic Problem for the Korteweg-de Vries Equation, *Funk. Analiz i Prilozh.* **8** (3) (1974), pp. 54–66 (in Russian)

[134] Oleinik O. A., Discontinuous Solutions of Nonlinear Equations, *Uspekhi Mat. Nauk* **12** (3) (1957), pp. 3–73 (in Russian).

[135] Ostrovskii L. A., Propagation of Wave Packets and Spatial Focussing in a Nonlinear Medium, *ZhETF* **51** (4) (1966), pp. 1189–1194 (in Russian).

[136] Pelinovskii E. N., *Nonlinear Dynamics of Tsunami Waves*, IPF AN SSSR, Gorkij, 1982 (in Russian).

[137] Popov M. M., A New Method of Computation of Wave Fields Using Gaussian Beams, *Wave Motion* **4** (1) (1982), pp. 85–97.

[138] Povsner A. J., Linear Method in Problems of Nonlinear Differential Equations with a Small Parameter, *Int. J. Nonlinear Mechanics* **9** (1974).

[139] Protas Yu. I., Semi-Classical Asymptotics of the Amplitude of Plane Wave Scattering on Non-Homogeneous Mediums, *Matem. Sbornik* **117** (4) (1982), pp. 495–515 (in Russian).

[140] Rockland C., Hypoellipticity on the Heisenberg Group-Representation-Theoretic Criteria, *Trans. Amer. Math. Soc.* **240** (1978), pp. 1–52.

[141] Rotschild L. P. and Stein E. M., Hypoelliptic Differential Operators and Nilpotent Groups, *Acta Math.* **137** (1976), pp. 247–320.

[142] Rotschild L. P., A Criterion for Hypoellipticity of Operators Contructed from Vector Field, *Comm. Part. Differ. Equat.* **4** (1979), pp. 645–699.

[143] Rusanov V. V. and Bezmonov I. V., On the Limit Profile of Nonlinear Discontinuity in Difference Schemes for the One-Dimensional Quasilinear Equation, Prepr. IPM, N69, Moscow, 1980 (in Russian).

[144] Scott A., *Active and Nonlinear Wave Propagation in Electronics*, Wiley-Interscience, New York, 1970.

[145] Semenov N. N., *On Some Problems of Chemical Kinetics and Reaction Ability*, Moscow, 1970 (in Russian).

[146] Serdyukova S. I., On Stability in the Uniform Metric of Systems of Difference Equations, *ZhVM i MF* **7** (3) (1967), pp. 497–509 (in Russian).

[147] Shubin M. A. *Pseudodifferential Operators and Spectral Theory*, Nauka, Moscow, 1978 (in Russian).

[148] Sjöstrand J. and Melin A., Fourier Integral Operators with Complex-Valued Phase Functions, *Proc. Int. Conf. in Nice on Fourier Int. Oper. 1974.*

[149] Souriau J.-M., Quantification Geometrique, *Comm. Math. Phys.* **1** (1966), pp. 374–398.

[150] Souriau J.-M., *Structure des systems dynamiques*, Dunod, Paris, 1970.

[151] Souriau J.-M., *Geométric sympletique et physique mathematique*, Paris, 1975.

[152] Souriau J.-M., Construction explicite de l'indice de Maslov. Applications, *Lect. Notes. Phys.* **50** (1976), pp. 117–148.

[153] Toda M., Vibration of Chain with Nonlinear Interaction, *J. Phys. Soc. Japan* **22** (1967), pp. 431–437.

[154] Treves F., Hypoelliptic Partial Differential Equations of Principal Type. Sufficient Conditions and Necessary Conditions, *Comm. Pure Appl. Math.* **24** (5) (1971), pp. 631–670.

[155] Trubowitz E., The Inverse Problem for Periodic Potentials, *Comm. Pure Appl. Math.* **30** (1977), pp. 321–337.

[156] Uhlmann G. A., Light Intensity Distribution in Conical Refraction, *Comm. Pure Appl. Math.* **35** (1) (1982), pp. 69–80.

[157] Varadhan S. R. S., Asymptotic Probabilities and Differential Equations, *Comm. Pure Appl. Math.* **19** (3) (1966), pp. 261–286.

[158] Varadhan S. R. S., On the Behaviour of the Fundamental Solution of the Heat Equation with Variable Coefficients, *Comm. Pure Appl. Math.* **20** (2) (1967), pp. 431–455.

[159] Vakulenko S. A. and Molotkov I. A., Waves in a Nonlinear Non-Homogeneous Medium, Concentrated in the Vicinity of the Given Curve, *DAN SSSR* **262** (3) (1982), pp. 587–591 (in Russian).

[160] Varchenko A. N., Newtonian Polyhedra and the Estimates of Oscillating Integrals, *Funk. Analiz* **10** (3) (1976), pp. 13–38 (in Russian).

[161] Vorob'ev Yu. M. and Dobrokhotov S. Yu., Semi-Classical Quantization of the Periodic Toda Chain from the Point of View of Lie Algebras, *TME* **54** (3) (1983), pp. 477–479 (in Russian).

[162] Weinberg B. R., *Asymptotical Method in Problems of Mathematical Physics*, MGU, Moscow, 1982 (in Russian).

[163] Weinstein A., On Maslov's Quantization Condition, *Lect. Notes Math.* **459** (1975), pp. 341–372.

[164] Weinstein A. I., Zhararov V. I., Novikov E. A., and Shifman M. A., Instanton Alphabet, *UFN* **134** (4) (1982) (in Russian).

[165] Whitham G. B., Nonlinear Dispersive Waves, *Proc. Roy. Soc.* **A283** (1393) (1965), pp. 283–291.

[166] Whitham G. B., A General Approach to Linear and Nonlinear Dispersive Waves Using a Lagrangian, *J. Fluid. Mech.* **22** (2) (1965), pp. 273–283.

[167] Whitham G. B., *Linear and Nonlinear Waves*, Wiley-Interscience, New York, 1974.

[168] Zabusky N. J., A Synergetic Approach to the Problem of Nonlinear Dispersive Wave Propagation and Interaction. In: W. F. Ames (ed.), *Proc. Sym. on Nonlinear Partial Differential Equations, Univ. of Delaware, Newark, 1965*, Academic Press, 1967, p. 223.

[169] Zeldovich Ya. B., Barenblatt G. I., Librovich V. V., and Makhviladze G. M., *Mathematical Theory of Burning and Explosion*, Nauka, Moscow, 1980 (in Russian).

[170] Zhikov V. V., Kozlov S. M., Oleinik O. A., and Kha T'en Ngoan, Averaging and *G*-Convergence of Differential Operators, *Uspekhi Mat. Nauk* **34** (5) (1979) (in Russian).
[171] Zinn-Justin J., Expansion around Instantons in Quantum Mechanics, *J. Math. Phys.* **22** (3) (1981), pp. 511–520.

Additional references

[172] Oleynik O. A., Panasenko G. P., and Iosifyan G. A., Asymptotical Expansion of Solutions of a System of Elasticity Theory with Rapidly Oscillating Coefficients, *DAN* **266** (1) (1982), pp. 18–22.
[173] Oleynik O. A., Panasenko G. P., and Iosifyan G. A., Asymptotical Expansion of Solutions of a System of Elasticity Theory in Perforated Domains, *Matem. sb.* **120** (1) (1983), pp. 22–41.
[174] Lions J. L., *Generalized Solutions of Hamilton–Jacobi Equations*, Pitman Advanced Publishing Program, Boston–London–Melbourne, 1982.
[175] Imanaliev M. I., *Asymptotical Methods in the Theory of Singularly Perturbated Integro-Differential Systems*, Ilim, Frunse, 1972.
[176] Danilov V. G. and Le Vu An, On Integral Fourier Operators, *Matem. sb.* **110** (152) (3) (1979), pp. 323–368.
[177] Melin A. and Sjöstrand J., Fourier Integral Operators with Complex-Valued Functions, *Lecture Notes Math.* **459** (1975), pp. 121–223.
[178] Mishchenko A. S., Sternin B. Yu., and Shatalov V. E., *Lagrangian Manifolds and the Method of Canonical Operator*, Nauka, Moscow, 1978.
[179] Maslov V. P. and Omel'yanov G. A., On Hugoniot Conditions for Infinitely Narrow Solutions of Simple Waves Equations, *Sib. Math. Journ.* **24** (2) (1983), pp. 172-182 (in Russian).

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

B. MAZUR

Modular Curves and Arithmetic

Introduction

The real number

$$\frac{\sin\frac{2\pi}{13}\cdot\sin\frac{5\pi}{13}\cdot\sin\frac{6\pi}{13}}{\sin\frac{\pi}{13}\cdot\sin\frac{3\pi}{13}\cdot\sin\frac{4\pi}{13}}$$

is a (fundamental) unit in the field $Q(\sqrt{13})$. It has been known since 1837 that one can systematically produce units in real quadratic fields (and hence produce solutions of Pell's equation) by trigonometric expressions of the above sort.[1]

More generally, the circular units, i.e., Norms of

$$\frac{1-e^{\frac{2\pi i a}{N}}}{1-e^{\frac{2\pi i}{N}}}$$

for relatively prime integers a and N, generate subgroups of finite index in the group of units of number fields F which are abelian over Q. The interest in circular units stems from the fact that they are "sufficiently many" explicitly constructed units, and also from the fact that their "position" in the full group of global units (and in certain groups of local units) bears upon deep arithmetic questions concerning F.

[1] The discoverer of this method [11] remarks that it is far less efficient for numerical calculation than is the method of continued fractions (a sentiment we may easily share by considering the displayed example above) and he rather envisions it "comme un rapprochement entre deux branches de la science des nombres".

By the work of [48], [49], [66]–[69], [60], we now have an analogous set-up for number fields which are abelian extensions of quadratic imaginary fields (the theory of *elliptic units*).

It has long been observed that there are resonances between the question of determining the group of units in a number field, and that of determining the group of rational points of an elliptic curve E over a number field. If K is a number field, the theorem of Mordell–Weil asserts that $E(K)$, the group of K-rational points of E, is finitely generated.

The problem of producing by some systematic means, sufficiently many explicitly constructed rational points in the Mordell–Weil group $E(K)$ seems, however, to be hampered from the outset by the fact that, in contrast to the problem of units, the rank of $E(K)$ is an erratic function of K about which we know little.

Nevertheless, thanks to recent work of Coates–Wiles, Rubin, Greenberg, Gross–Zagier and Rohrlich, a fascinating picture is emerging in the study of the Mordell–Weil group of elliptic curves, which reminds one of the already established theory of elliptic units. The general problem encompassing this recent work is that of understanding, for E/Q an elliptic curve over Q,[2] the behavior of the Mordell–Weil group $E(L)$ where L is a (varying) abelian extension of a (fixed) quadratic imaginary number field. In these Mordell–Weil groups one can produce an "orderly" and systematic supply of rational points (*Heegner points*). The Heegner points do not always account for the full rank of the Mordell–Weil group, but, in some statistical sense, yet to be made precise, they may very well provide the major contribution (the "*dominant term*") whose asymptotic nature is succinctly describable. In consequence, the erratic nature of the rank of $E(L)$ for varying L would be due to the presence of an "error term" whose fluctuations would be, perhaps, truly difficult to come to grips with, but nevertheless comparatively minor in amplitude.

Since it is usually sound mathematical practice to deal with dominant terms, before getting down to error terms, this deserves close study. What makes this study even more attractive is the recent development of new tools which might be brought to bear on these problems (p-adic heights; two-variable p-adic L functions, and a number of novel approaches to the handling of special values of classical L functions).

The object of this expository paper is to sketch the "emerging picture" much of which is still conjectural, and to describe recent advances towards its development.

[2] Which is assumed to admit a parametrization by modular functions.

I. Arithmetic

1. Weil parametrizations. The Weil–Taniyama conjecture asserts that any elliptic curve over $\boldsymbol{Q}$ can be parametrized by one of the family of modular curves $X_0(N)$, for $N = 1, 2, \ldots$ Recall that the modular curve $X_0(N)/\boldsymbol{Q}$ is Shimura's canonical model of the Riemann surface $\mathscr{H}/\Gamma_0(N)$, the quotient of the upper half-plane by the action of the subgroup $\Gamma_0(N)$ consisting of matrices $\begin{bmatrix} ab \\ cd \end{bmatrix}$ in $\mathrm{SL}_2(\boldsymbol{Z})$ with $c \equiv 0 \bmod N$.

It has been abundantly clear for years that one has a much more tenacious hold on the arithmetic of an elliptic curve $E/\boldsymbol{Q}$ if one supposes that it is so parametrized. *We will suppose so*, and, moreover, with no loss of generality, we suppose that we have a *Weil parametrization*: a non-trivial morphism φ: $X_0(N) \to E$ defined over $\boldsymbol{Q}$ which brings the holomorphic differential on E to a new form, and which brings the cusp ∞ to the origin in E. The integer N for which this holds is an isogeny-class invariant of E, called the *conductor* of E. See [40] for tables listing quantities of elliptic curves possessing Weil parametrizations. Any complex multiplication elliptic curve over $\boldsymbol{Q}$ admits a Weil parametrization.

Remark. If E admits a Weil parametrization by $X_0(N)$ and if d is a divisor of N which is the product of an even number of distinct primes, and is relatively prime to N/d, then by the work of Ribet and Jacquet-Langlands, E also admits a parametrization by a certain "Shimura curve" attached to the quaternion algebra over $\boldsymbol{Q}$ which is nonsplit at the prime divisors of d.

For a more comprehensive study of "Heegner points" (cf. Section 4 below) on E, it may be useful to deal with the entire assortment of Shimura curve parametrizations of E, and not only its Weil parametrization.

2. Mordell–Weil rank. Let L/K be a finite abelian extension of number fields and E/K and elliptic curve over K. By the Mordell–Weil theorem $E(L)$ is finitely generated. We may decompose the complex vector space $E(L) \otimes \boldsymbol{C}$ as a direct sum:

$$E(L) \otimes \boldsymbol{C} = \bigoplus_{\chi} V(E, \chi)$$

where χ: $\mathrm{Gal}(L/K) \to \boldsymbol{C}$ [3] ranges through all characters, and $V(E, \chi)$ is the χ-eigensubspace in $E(L) \otimes \boldsymbol{C}$. The classical canonical normalized

[3] For an exposition of this theory, see Chapter IV of [34].

height pairing $\langle\,,\rangle$ gives $E(L)\otimes C$ and each $V(E,\chi)$ a nondegenerate Hermitian structure. Let $r(E,L)$ denote the rank of $E(L)$ and $r(E,\chi)$ the complex dimension of $V(E,\chi)$.

One has hardly begun to achieve significant understanding of the asymptotics of the function

$$\chi \mapsto r(E,\chi)$$

for fixed E/K. Two restricted types of "variation of χ" come to mind: *horizontal variation*, where one allows χ to run through all characters of fixed order, and *vertical variation*, where one considers all characters whose conductors have prime divisors belonging to a fixed, finite, set of primes S.

In a series of papers ([14]–[16]), Goldfeld and co-authors have considered the problem of horizontal variation for quadratic characters χ, where $K = \boldsymbol{Q}$. Goldfeld conjectures that the "average value of $r(E,\chi)$" is 1/2. This is consistent with the partial results he has obtained, the most striking being his theorem with Hoffstein and Patterson which implies that, when E/Q has complex multiplication, there are an infinity of quadratic χ over $\boldsymbol{Q}$ such that $r(E,\chi) = 0$.[4] The conjecture is also in accord with the qualitative results that are being obtained in the case of "vertical variation"; namely, that the bulk of the contribution to the Mordell–Weil rank over abelian extensions of quadratic number fields is of a size that would be given by the most conservative estimate compatible with the parity restrictions forced by the functional equation of associated L functions, via the conjectures of Birch and Swinnerton-Dyer. It would be interesting to experiment further with "horizontal variation" (e.g., cubic characters) to get a broader view.

3. Anti-cyclotomic extensions. Let $k\subset \boldsymbol{C}$ be a quadratic imaginary field. An *anti-cyclotomic extension* of k is a finite abelian extension L/k such that $L/\boldsymbol{Q}$ is a Galois extension, whose Galois group is a generalized dihedral group in the following sense: there is an involution σ of L which induces complex conjugation on k, and such that for every $g\in \mathrm{Gal}(L/k)$ we have $\sigma g\sigma^{-1} = g^{-1}$. An *anti-cyclotomic character* of k is a continuous homomorphism of $\mathrm{Gal}(\bar{k}/k)$ to $\boldsymbol{C}^*$ whose $k/\boldsymbol{Q}$ conjugate is equal to its inverse. Thus

[4] For more general results concerning the nonvanishing of the L function of modular forms twisted by quadratic characters evaluated at the center of the critical strip, see § 9 below, Waldspurger [63], and the exposition of Waldspurger's results in [62].

the anti-cyclotomic extensions of k are precisely those abelian extensions of k all of whose associated characters are anti-cyclotomic.

Examples of anti-cyclotomic extensions are given by ring-class fields over k: Recall that for every positive number c, the *order of conductor* c, $\mathcal{O}_c$, in the full ring of integers $\mathcal{O}$ of k is the module $\boldsymbol{Z}+c\cdot\mathcal{O}$, immediately seen to be a subring of $\mathcal{O}$ of finite index. Viewing $\mathcal{O}_c \subset \mathcal{O} \subset k \subset \boldsymbol{C}$ as a lattice in the complex plane, and viewing the elliptic modular function j as a function of lattices, one shows that $j(\mathcal{O}_c)$ is a real algebraic number. One defines H_c, the ring class field of k of conductor c, to be $k(j(\mathcal{O}_c)) \subset \boldsymbol{C}$. Class field theory establishes an isomorphism of groups

$$\mathrm{Gal}(H_c/k) \underset{\cong}{\leftarrow} \mathrm{Pic}(\mathcal{O}_c)$$

where $\mathrm{Pic}(\mathcal{O}_c)$ denotes the group of invertible $\mathcal{O}_c$-ideal classes. Taking our involution σ to be the automorphism of H_c induced by complex conjugation, one sees that H_c is an anti-cyclotomic extension of k. Moreover, any anti-cyclotomic extension of k is contained in a ring-class field H_c for some c. If χ is an anti-cyclotomic character of k, and c is the minimal integer for which χ belongs to H_c then χ is said to be *primitive* on H_c.

A character χ on $\mathrm{Gal}(\bar{\boldsymbol{Q}}/\boldsymbol{Q})$ is anti-cyclotomic when restricted to $\mathrm{Gal}(\bar{k}/k)$ if and only if χ is quadratic.

4. Heegner points. These points were first discovered by Heegner [25] and developed by Birch [4], [5]. They may be now seen as a particular case of a quite general construction ("special points and cycles") on Shimura varieties. Heegner points call attention to themselves by virtue of their being a plentiful and orderly supply of points on modular curves $X_0(N)$ rational over anti-cyclotomic extensions of quadratic number fields. Given a Weil parametrization $\varphi\colon X_0(N)\to E$, the image of the Heegner points will generate some portion of the Mordell–Weil group of E. The marvelous result of Gross–Zagier gives us some idea of the portion generated. We sketch the definition of Heegner points.[5]

If K is a field of characteristic 0, a pair $(\mathcal{E}, C_N)$ consisting of an elliptic curve $\mathcal{E}$ over K, and a cyclic subgroup of order N, $C_N \subset \mathcal{E}$, rational over K defines a K-valued point of $X_0(N)$, which we shall denote $j(\mathcal{E}, C_N)$.

---

[5] We restrict ourselves here to Heegner points on E arising from its Weil parametrization. This accounts for the restrictedness of the Heegner Hypothesis below.

As hinted in Section 1, it will be eventually necessary to consider the Heegner points of E arising from its various Shimura curve parametrization as well. For this theory, see [28], especially Chapter 3.

Now let k denote a quadratic imaginary field contained in C, $\mathcal{O}_c$ an order of conductor $c > 0$ contained in the ring of integers of k.

Heegner Hypothesis. We say that the pair $(\mathcal{O}_c, N)$ satisfies the *Heegner Hypothesis* if there exists an invertible $\mathcal{O}_c$-ideal $\mathcal{N} \subset \mathcal{O}_c$ such that $\mathcal{O}_c/\mathcal{N}$ is cyclic of order N.

This is the case when, for example, N is prime to the discrimant of $\mathcal{O}_c$, and every prime dividing N splits in the ring of integers of k. Suppose now, that $(\mathcal{O}_c, \mathcal{N})$ satisfies the Heegner hypothesis and choose an ideal $\mathcal{N}$ as above.

Let $y(\mathcal{O}_c, \mathcal{N})$ denote the point $j(C/\mathcal{O}_c, \mathcal{N}^{-1}/\mathcal{O}_c)$ in $X_0(N)(C)$. One learns from the classical theory of complex multiplication that $y(\mathcal{O}_c, \mathcal{N})$ is defined over the field $H_c \subset C$, and its full set of Galois-conjugates over k is operated on principally and transitively by $\mathrm{Gal}(H_c/k)$. We refer to $y(\mathcal{O}_c, \mathcal{N})$ or any of its Galois-conjugates as basic Heegner points (*of type* $(\mathcal{O}_c, \mathcal{N})$).

Let $x(\mathcal{O}_c, \mathcal{N})$ denote the image of $y(\mathcal{O}_c, \mathcal{N})$ under the Weil parametrization φ, in the Mordell–Weil group of E over H_c. We refer to the $x(\mathcal{O}_c, \mathcal{N})$ and their Galois-conjugates as basic Heegner points on E.

If χ is a primitive character on $\mathrm{Gal}(H_c/k)$, define

$$x(\chi, \mathcal{N}) = \sum \chi^{-1}(g)\cdot x(\mathcal{O}_c, \mathcal{N})^g \in V(\chi).$$

One easily sees that, up to sign, $x(\chi, \mathcal{N})$ is independent of the choice of $\mathcal{N}$, and hence depends only upon χ.

We shall say that the pair (k, E) satisfies the *Heegner hypothesis* (*for the Weil parametrization*) if the pair $(\mathcal{O}, N)$ satisfies the Heegner hypothesis, as above.

II. The (C-valued) analytic theory

5. L functions. Let E be a Weil curve, i.e., E is an elliptic curve admitting a parametrization by $X_0(N)$ over Q as in Section 1.

Let k be a quadratic imaginary field, χ a finite Hecke character for k of conductor $\mathfrak{f}$ with values in C. If v is a valuation of k, let χ_v denote the v-adic component of χ.

The *Euler factor* $L_v(E/k_v, \chi_v, s) = L_v$ is defined for places $v \nmid N\cdot\mathfrak{f}\cdot\infty$ by the formula:

$$L_v^{-1} = 1 - a_v\cdot\chi_v(\pi_v)Nv^{-s} + \chi_v^2(\pi_v)Nv^{1-2s}$$

where N_v is the cardinality of $k(v)$, the residue field of v, π_v is a v-adic uniformizer, and a_v is that integer such that $1+Nv-a_v = \#\big(E\big(k(v)\big)\big)$, the number of rational points of the reduction of E to $k(v)$.

For the remaining non-archimedean primes, L_v can be given explicitly as a polynomial in Nv^{-s} and for the archimedean prime, $L_v = (2\pi)^{-2s}\Gamma(s)^2$, where $\Gamma(s)$ is the classical Γ-function.

One defines the (classical) L-function $L(E, \chi, s)$ to be the product of all the Euler factors L_v (including $v = \infty$). It is easily seen to be convergent to an analytic function in a suitable right half-plane. By Rankin's method (cf. [26]), one knows that $L(E, \chi, s)$ extends to an entire function satisfying the functional equation:

$$L(E, \chi, 2-s) = e\cdot A^{s-1}\cdot L(E, \chi^{-1}, s)$$

for $e = e(E, \chi) \in C$ and $A = A(E, \chi)$, suitable constants.

Let us distinguish two cases:

The Exceptional Case. E/Q has complex multiplication by the field k.

The Generic Case. Either E/Q has no complex multiplication, or its field of complex multiplication is different from k.

In the exceptional case, the L function $L(E, \chi, s)$ factors into a product of two L functions attached to Grossencharacters over k. Specifically,

$$L(E, \chi, s) = L(\Phi, \chi, s)\cdot L(\Phi\varepsilon, \chi, s),$$

where Φ is the Grossencharacter attached to E/Q and ε is the quadratic character belonging to k. Moreover, one has the functional equation:

$$L(\Phi, \chi, 2-s) = e\cdot A^{s-1}L(\Phi, \chi^{-1}, s)$$

for suitable constants $e = e(\Phi, \chi)$, $A = A(\Phi, \chi)$.

For anti-cyclotomic characters, the functional equation simplifies to read:

$$L(E, \chi, 2-s) = e\cdot A^{s-1}\cdot L(E, \chi, s)$$

and

$$L(\Phi, \chi, 2-s) = e\cdot A^{s-1}\cdot L(\Phi, \chi, s) \quad \text{(in the exceptional case)},$$

where, in either case, the constant e is ± 1.

6. Signs. If χ is anti-cyclotomic, define the *sign* of (E, χ) to be $e(\varphi, \chi)$ in the exceptional case and $e(E, \chi)$ in the generic case. Thus the sign of (E, χ) determines the parity of the order of vanishing of $L(E, \chi, s)$ at

$s = 1$ in the generic case, and the parity of one-half that order in the exceptional case.

The problem of calculating sign (E, χ) has been studied extensively. From a representation-theoretic point of view, we have Jacquet's [26], Chapter V, which draws on the techniques of [27]. See also Weil's [65]. From a slightly different point of view, see Kurcanov [32]. For calculations germane to our setting, see Gross [20] for the generic case and for the exceptional case, Greenberg [17]. For the interesting question of signs of elliptic curves over $\mathbf{Q}$, see Kramer and Tunnell [33].

For later purposes we content ourselves with two formulas valid in the generic case:

(a) Let N be prime to the discriminant of k. Let χ_0 be the principal Hecke character over k, and ε the quadratic Dirichlet character attached to k. Then $\operatorname{sign}(E, \chi_0) = -\varepsilon(N)$ (e.g., if the pair (k, E) satisfies the Heegner hypothesis of Section 5, then $\operatorname{sign}(E, \chi_0) = -1$).

(b) If χ is an anti-cyclotomic character of conductor prime to N, then $\operatorname{sign}(E, \chi) = \chi(N) \cdot \operatorname{sign}(E, \chi_0)$.

7. Analytic rank versus arithmetic rank. Let $\varrho(E, \chi)$ denote the order of vanishing of $L(E, \chi, s)$ at $s = 1$ (the "analytic rank"). Viewing χ as a Galois character via class field theory, we have the "arithmetic rank" $r(E, \chi)$ defined as in Section 2.

The conjectures of Birch and Swinnerton-Dyer (weakened and strengthened a bit) lead one to

The Rank Conjecture: $\varrho(E, \chi) = r(E, \chi)$.

8. The theory of Gross–Zagier. Gross and Zagier [22], [23] establish a magnificent formula which is, as they describe it, a new kind of Kronecker limit formula. A version of (special cases of) this formula had been previously conjectured by Birch [4], [5] and Stephens with significant numerical evidence compiled in its support. We shall describe, below, a slightly weakened version of the Gross–Zagier theorem.

Let

$$f(z)\,dz = \sum_{n\geq 1} a_n e^{2\pi i n z}\,dz$$

(with a_1 equal to 1 and not $2\pi i$) be the new form on $\Gamma_0(N)$ associated to E, in the sense that there is a holomorphic differential ω on E, such that $\varphi^*\omega = f(z)\,dz$.

Let χ be an anti-cyclotomic character on k, primitive on H_c. *Suppose that c is relatively prime to N, and the Heegner hypothesis holds for $(\mathcal{O}, N)$.*

Then the *Heegner hypothesis* also holds for $(\mathcal{O}_c, \mathcal{N})$. Let $x(\chi, \mathcal{N})$ be the Heegner point in $V(E, \chi)$, the χ-part of the Mordell–Weil group as in § 2. Let

$$\|x(\chi, \mathcal{N})\|^2 = \langle x(\chi, \mathcal{N}), x(\chi, \mathcal{N})\rangle$$

be the square of the norm of $x(\chi, \mathcal{N})$ in the Hermitian structure of $V(E, \chi)$ determined by the height pairing. This is independent of the choice of $\mathcal{N}$.

Under our hypotheses, we are in the generic case and the sign of (E, χ) is -1. Therefore $L(E, \chi, 1)$ vanishes.

THEOREM (Gross–Zagier). *There is a positive constant C such that*

$$L'(E, \chi, 1) = C \cdot \|x(\chi, \mathcal{N})\|^2.$$[6]

The reader is referred to the forthcoming papers of Gross and Zagier for a full account. One can read the formula in either direction, with important consequences ensuing. Notably, by finding Heegner points which vanish, Gross and Zagier manage to produce L-functions with high order of vanishing at $s = 1$, which by the prior work of Goldfeld [13] then provides an effective solution to the problem of listing quadratic imaginary fields of class number $\leqslant B$ for any bound B.

But also, in the context in which their theorem applies, if $L'(E, \chi, 1)$ doesn't vanish, their formula shows that the Heegner point $x(\chi, N)$ is *nonzero* and hence if $\varrho(E, \chi) = 1$, then $r(E, \chi) \geqslant 1$.

Gross and Zagier prove their formula by first providing just the right (infinite sum) expression for each side of their formula. These expressions have the virtue that there is a one:one correspondence between terms on the left and on the right. To prove their formula, they must establish an equality between corresponding terms and then deal with the highly nontrivial convergence problems that stand in their way. The infinite sum expression on the left comes naturally from Rankin's method for treating $L(E, \chi, s)$. The infinite sum expression on the right comes from an expression for the archimedean contribution to the height in terms of Green's functions as well as a finite sum coming from the nonarchimedean contributions.

9. Non vanishing of $\varrho(E, \chi)$: the theory of Waldspurger. The theory alluded to has a number of deep arithmetic consequences. See Waldspurger's

[6] Gross and Zagier give an expression for C in terms of the Petersson inner product of f with itself, and elementary invariants. They can also work with certain Shimura curve parametrizations in place of the Weil parametrization to obtain a formula for the canonical heights of "Heegner points" coming from these parametrizations as well.

account of this theory in the proceedings of this Congress. One corollary of his work is:

THEOREM. *There is an infinity of odd quadratic Dirichlet characters* χ *such that if* k_χ *is the (quadratic imaginary) field associated to* χ, *then* (k_χ, N) *satisfies the Heegner hypothesis and* $\varrho(E/Q, \chi) = 0$.

10. Evidence for the "rank conjecture". By the fundamental work of Coates–Wiles [8],[7] as refined by Karl Rubin [54], we know that *if* E *has complex multiplication, then* $\varrho(E, \chi) = 0$ *implies that* $r(E, \chi) = 0$.

The result of Gross and Zagier tells us that *if* E/Q *is an arbitrary (Weil parametrized) elliptic curve, the Heegner hypothesis is satisfied for* $(\mathcal{O}_c, N)$, *and* χ *is an anti-cyclotomic character primitive on* H_c, *then,* $\varrho(E, \chi) = 1$ *implies that* $r(E, \chi) \geqslant 1$. If we are in the "exceptional case", the Heegner hypothesis is not met. But by choosing an auxiliary quadratic imaginary field, using results of Waldspurger § 9, [62], [63] one can show that $\varrho(E, \chi) = 2$ implies that $r(E, \chi) \geqslant 2$.

11. Galois-conjugate characters. If $\chi\colon \mathrm{Gal}(\bar{k}/k) \to C^*$ is a character, let $Q(\chi)$ denote the subfield of C generated by its values. If $\tau\colon Q(\chi) \to Q(\chi)$ is a field automorphism and χ^τ the composition of τ with χ, we refer to χ and χ^τ as *Galois-conjugate characters*. Since $r(E, \chi) = r(E, \chi^\tau)$, the "rank conjecture" would imply that

CONJECTURE.[8]

$$\varrho(E, \chi) = \varrho(E, \chi^\tau)$$

and for even this weaker conjecture we have only fragmentary evidence. Namely: By some results of Shimura, or by the theory of modular symbols, one knows that $\varrho(E, \chi)$ *vanishes if and only if* $\varrho(E, \chi^\tau)$ *does.*

It follows from the theorem of Gross and Zagier that *if* χ *is anti-cyclotomic, primitive on* H_c, c *is prime to* N *and the Heegner hypothesis holds for* $(\mathcal{O}, N)$, *then* $\varrho(E, \chi) = 1$ *if and only if* $\varrho(E, \chi^\tau) = 1$, *and in the exceptional case,* $\varrho(\Phi, \chi) = 1$ *if and only if* $\varrho(\Phi, \chi^\tau) = 1$.

12. Vertical anti-cyclotomic variation and Greenberg's theory. In his beautiful paper [17] (see also the subsequent reference [18]), Ralph Greenberg proved the first deep general theorems about the vertical anti-cyclotomic variation of Mordell–Weil rank. Greenberg works in the exceptional case,

[7] See also [21], [59].

[8] Compare Conjecture 2.7 of [10].

and to prove his theorems he initiated an elegant method which blends archimedean techniques (for estimating special values of L functions attached to Grossencharacters) with p-adic techniques (where Greenberg makes essential use of Yager's two-variable p-adic L functions for more than one prime p and of Perrin–Riou's descent-theoretic study of the p-Selmer group [46]).

Subsequently, Rohrlich [50] discovered another approach to the same vertical problem which has very surprising points of similarity and difference to Greenberg's theory. For example, the archimedean estimate proved by Greenberg is a certain Abel mean summability property for the special values $L(\Phi^{2k+1}, \chi, k+1)$. Greenberg shows that under the appropriate condition of sign, the Abel sums taken over k in certain arithmetic progressions are nonzero. To do this, he makes use of the (archimedean) Roth's theorem. He then appeals to the theory of p-adic L functions, and deduces parts A of the theorem and corollary below as well as a weaker version of part B of the corollary $\left(r_p(E, \chi) \geqslant 2\right)$.

Rohrlich, on the other hand, establishes a different type of convergence. He makes no use of p-adic L functions, but does use the (p-adic) version of Roth's theorem in his theory. Is there a way of unifying the two approaches?

13. Rohrlich's theory of Galois-averages. Let us suppose that we are in the exceptional case, with Grossencharacter Φ. Fix a finite set of primes S of k, and let X_S denote the set of anti-cyclotomic characters over k whose conductor has prime divisors belonging to S. Then $X_S = X_S^+ \coprod X_S^-$, where $X_S^\pm$ denotes the subset of characters χ for which the sign of (E, χ) is ± 1. Let

$$\mathscr{L}(\Phi, \{\chi\}) = 1/[\boldsymbol{Q}(\chi) \colon \boldsymbol{Q}] \cdot \sum_{\tau \in \mathrm{Gal}(\boldsymbol{Q}(\chi)/\boldsymbol{Q})} L(\Phi, \chi^\tau, 1)$$

be the average value of $L(\Phi, \chi^\tau, 1)$ over the set of characters χ^τ which are Galois-conjugate to χ. Rohrlich's remarkable discovery is that *the average values $\mathscr{L}(\Phi, \{\chi\})$ tend to a nonzero limit as χ_i ranges through an infinite set of characters in X_S^+ such that $f(\chi_i)$ divides $f(\chi_{i+1})$.*

Letting

$$\mathscr{L}'(\Phi, \{\chi\}) = 1/[\boldsymbol{Q}(\chi) \colon \boldsymbol{Q}] \cdot \sum_{\tau \in \mathrm{Gal}(\boldsymbol{Q}(\chi)/\boldsymbol{Q})} L'(\Phi, \chi^\tau, 1)$$

denote the average value of the first derivatives, Rohrlich shows that for any constant $c \in \boldsymbol{R}$ there are only a finite number of χ in X_S^- such that $|\mathscr{L}'(\varphi, \{\chi\})| \leqslant c$.

Using known facts (as described in Section 11) concerning invariance of $\varrho(E, \chi)$ under Galois conjugation of χ, Rohrlich deduces:

THEOREM. *Suppose that we are in the exceptional case. Then for all but a finite number of characters χ in X_S,*

$$\varrho(E, \chi) = 0 \quad \textit{if the sign of } (E, \chi) \textit{ is } +1 \tag{A}$$

and

$$\varrho(E, \chi) = 2 \quad \textit{if the sign of } (E, \chi) \textit{ is } -1. \tag{B}$$

Using known facts (as described in Section 10), he then obtains:

COROLLARY. *Suppose that we are in the exceptional case. Then for all but a finite number of characters χ in X_S,*

$$r(E, \chi) = 0 \quad \textit{if the sign of } (E, \chi) \textit{ is } +1 \tag{A}$$

and

$$r(E, \chi) \geqslant 2 \quad \textit{if the sign of } (E, \chi) \textit{ is } -1. \tag{B}$$

Briefly, the technique of Rohrlich's proof is to express the Dirichlet series $L(\Phi, \chi_i, 1)$ as a sum of two types of terms: those terms that are visibly invariant under Galois-conjugation of χ_i (the "fixed" part) and those that "vary" (the "varying part"). The fixed part can be expressed as a nonzero multiple of $L(\varepsilon, 1)$, that multiple being constant for large enough $f(\chi_i)$. The varying part is shown to have Galois-average tending to zero, by an ingenious use of (the p-adic version of) Roth's theorem. His treatment of $L'(\Phi, \chi_i, 1)$ is similar.

In the light of some work of Asai [1], [2], it is tempting to conjecture that the format of Rohrlich's theory carries over in the generic case as well (although the techniques of proof may not).

Let

$$\mathscr{L}(E, \{\chi\}) = 1/[Q(\chi): Q] \cdot \sum_{\tau \in \mathrm{Gal}(Q(\chi)/Q)} L(E, \chi^{\tau}, 1)$$

and

$$\mathscr{L}'(E, \{\chi\}) = 1/[Q(\chi): Q] \cdot \sum_{\tau \in \mathrm{Gal}(Q(\chi)/Q)} L'(E, \chi^{\tau}, 1).$$

CONJECTURE. *Suppose that we are in the generic case.*
Let χ_i be a sequence of characters in X_S^+ such that $f(\chi_i)$ divides $f(\chi_{i+1})$.

Then

$$\mathscr{L}(E, \{\chi_i\}) \quad \textit{tends to a nonzero limit.}$$

Let χ_i be an infinite sequence of characters in X_S^-. Then

$$|\mathscr{L}'(E, \{\chi_i\})| \textit{ tends to infinity.}$$

A consequence of this conjecture together with Rohrlich's theorem is the following:

CONJECTURE. *For all but a finite number of characters in X_S we have*

$$\varrho(E, \chi) = 0 \quad \textit{if the sign of } (E, \chi) \textit{ is } +1,$$

$$\varrho(E, \chi) = 1 \quad \textit{if the sign of } (E, \chi) \textit{ is } -1, \textit{ and we are in the generic case,}$$

$$\varrho(E, \chi) = 2 \quad \textit{if the sign of } (E, \chi) \textit{ is } -1, \textit{ and we are in the exceptional case.}$$

14. Cyclotomic vertical variation. If E/Q is a Weil parametrized elliptic curve, S a finite set of primes of Q and $Y_S =$ the set of Dirichlet characters (over Q) whose conductors have prime divisors belonging to S, is it true that there are no more than a finite number of characters $\chi \in Y_S$ such that $\varrho(E, \chi) \neq 0$?

Rohrlich has recently proved this to be true under the hypothesis that no prime in S divides N, the conductor of E. He adapts his method of Galois-averages to obtain this result. Using it, together with the work of Rubin [54] complementing that of Coates–Wiles [8], he obtains

THEOREM. *Let E/Q be an elliptic curve of complex multiplication and conductor equal to N. Let F be an (infinite) abelian extension of Q unramified at ∞ and outside a finite set of primes S, none of which divide N. Then $E(F)$ is finitely generated.*

The above theorem would also be true for arbitrary Weil curves E/Q if, for example, the rank conjecture (§ 7) were true. The reader is also referred to the papers [55] and [56], where Rubin and Wiles prove something close to the above theorem by first showing certain mod p congruences between special values of L functions and Bernoulli numbers, and then invoking the deep result of Washington [64] and their extension by Friedman [12] to show that the Bernoulli numbers in question are rarely congruent to zero mod p. For a general discussion on congruences, see Stevens [61].

15. Z_p-extensions and p-adic logarithms over k. If K is a numberfield and A_K^*/K^* its idele class group, then any nontrivial continuous homomorphism from A_K^*/K^* to R^+, the additive group of reals, must factor through the idele norm mapping $A_K^*/K^* \to R^*/(\pm 1)$. Consequently, to give such a homomorphism is equivalent to giving a "logarithm mapping", i.e., an isomorphism $R^*/(\pm 1) \underset{\cong}{\to} R^+$.

Define a *p-adic logarithmic character over* K (or a *p-adic logarithm over* K for short) to be any continuous homomorphism

$$\lambda\colon A_K^*/K^* \to Q_p^+.$$

The space of p-adic logarithms over K forms a finite-dimensional Q_p-vector space of dimension $\geq r_2+1$, where r_2 is the number of complex archimedean places of K. A conjecture of Leopoldt asserts that its dimension is precisely r_2+1; this is known to be true for fields K which are abelian over Q, by the work of Brumer.

In the case where $K = Q$, the space of p-adic logarithms is one-dimensional, with a natural choice of generator σ determined by the prescription that σ restricted to $Q_p^* \hookrightarrow A_K^*$ be $\log_p$, the standard p-adic logarithm. We refer to σ as the *p-cyclotomic logarithm over* Q; for K any number field, the *p-cyclotomic logarithm over* K is the composition of σ with the norm mapping $A_K^*/K^* \to A_Q^*/Q^*$.

By Class field theory, any p-adic logarithm λ over K determines (and is determined by) a unique continuous homomorphism

$$\lambda_{\mathrm{Gal}}\colon \mathrm{Gal}(\bar{K}/K) \to Q_p^+.$$

If λ is nontrivial, then the image of λ_{Gal} is $p^N \cdot Z_p$ for some integer N, and so the kernel of λ_{Gal} has, as fixed field, a Z_p-extension L/K, i.e., a Galois extension such that $\mathrm{Gal}(L/K)$ is isomorphic to the additive group Z_p. We refer to L/K as the Z_p-extension of K *cut out by* λ. The space of Z_p-extensions of K are then in one:one correspondence with the Q_p-projective space of lines through the origin in the vector space of p-adic logarithms over K.

If k is a quadratic imaginary number field, by an *anti-cyclotomic p-adic logarithm* we mean a nontrivial logarithm over k whose k/Q conjugate is equal to its negative. *Choose such an anti-cyclotomic logarithm* τ. Then any p-adic logarithm of k is a Q_p-linear combination of σ and τ.

The theory of Z_p-extensions was initiated by Iwasawa as a suitable framework for the study of asymptotic questions concerning the rate of growth of (p-primary components of) ideal class groups. For the arithmetic of elliptic curves, as John Coates once remarked, Z_p-extensions serve as an invaluable tool: they provide an amenable setting for the study of the p-adic vertical variation problem (e.g., as described in Section 2), a setting which is congenial to the method of p-power descent. Moreover, via their associated p-adic logarithms, $\mathbf{Z}_p$-extensions give rise to a theory of p-adic heights (for p ordinary) and to a theory of p-adic L functions which, conjecturally, should mesh very well with the arithmetic theory and provide a manageable computational handle on it.

16. The pro-p-Selmer group. If one is to make use of p-power descent theory to study the Mordell–Weil group, it is natural to work with the $\mathbf{Z}_p$-module

$$E(K)_p \underset{\text{defn}}{=} [E(K)/\text{torsion}]\otimes \mathbf{Z}_p$$

rather than with $E(K)$ itself. However, even this $\mathbf{Z}_p$-module is only indirectly approachable by descent-theoretic methods; for example, with our present state of knowledge, it would be difficult to prove any regularity of the growth of the $\mathbf{Z}_p$-rank of $E(K_n)_p$ as K_n runs through the n^{th} layers of a $\mathbf{Z}_p$-extension L/K.

It has been traditional to replace $E(K)_p$ by the pro-p-Selmer group $S_p(E/K)$ for which such "growth regularity" theorems can be proved, and which is conjecturally (via the Shafarevitch–Tate conjecture) equal to $E(K)_p$. We sketch the definition of the pro-p-Selmer group.

Let K be any number field, E an elliptic curve over K and p a prime number. Recall the classical p-Selmer group, defined to make the square

$$\begin{array}{ccc} \text{Selmer}_p(E/K) & \hookrightarrow & H^1(\text{Gal}(\bar{K}/K), E[p^\infty]) \\ \downarrow i & & \downarrow j \\ Ш(E/K)[p^\infty] & \hookrightarrow & H^1(\text{Gal}(\bar{K}/K), E[p^\infty]) \end{array}$$

cartesian, where i and j are the natural morphisms, and $[p^\infty]$ means the union of the kernels of multiplication by p^n for all $n \geqslant 0$. The Shafarevitch–Tate group of E over K is denoted $Ш(E/K)$. It is defined to be the intersection of the kernels

$$H^1(\text{Gal}(\bar{K}/K), E) \to H^1(\text{Gal}(\bar{K}_v/K_v), E),$$

where v runs through all places (archimedean and nonarchimedean) of K. There is a natural mapping

$$E(K)\otimes \mathbf{Q}_p/\mathbf{Z}_p \hookrightarrow \text{Selmer}_p(E/K)$$

which is injective, and induces an injection on Tate modules. Note that

$$E(K)_p = T_p(E(K)\otimes \mathbf{Q}_p/\mathbf{Z}_p) = \text{Hom}(\mathbf{Q}_p/\mathbf{Z}_p, E(K)\otimes \mathbf{Q}_p/\mathbf{Z}_p).$$

Let $S_p(E, K)$ stand for

$$T_p(\mathrm{Selmer}_p(E/K)) = \mathrm{Hom}(\boldsymbol{Q}_p/\boldsymbol{Z}_p, \mathrm{Selmer}_p(E/K)).$$

The Shafarevitch–Tate conjecture implies that

CONJECTURE. *The natural injection*

$$E(K)_p \hookrightarrow S_p(E/K)$$

is an equality.

If $r_p(E, K)$ denotes the $\boldsymbol{Z}_p$-rank of $S_p(E/K)$, we have the inequality

$$r(E, K) \leqslant r_p(E, K)$$

(conjecturally an equality).

17. $\boldsymbol{Z}_p$-extensions and universal norms. Let K be a number field L/K, an infinite Galois extension with Galois group Γ isomorphic to $\boldsymbol{Z}_p$. Let K_n be the "n-th layer" of the extension L/K with Galois group Γ_n isomorphic to $\boldsymbol{Z}/p^n\boldsymbol{Z}$. Denote by Λ the Iwasawa ring

$$\boldsymbol{Z}_p[[\Gamma]] = \varprojlim_n \boldsymbol{Z}_p[\Gamma_n].$$

By a system of Γ_n-modules, $A_n\,(n \geqslant 0)$ is meant a sequence of such modules together with inclusions

$$i\colon A_j \to A_k, \quad j \leqslant k$$

and "norm" mappings

$$N_{k,j} = N\colon A_k \to A_j, \quad k \geqslant j$$

compatible with Γ-module structures such that the composition $iN\colon A_k \to A_k$ is given by the natural norm mapping[9] (from Γ_k to Γ_j).

Given such a system, define *the subsystem of a universal norms* by the rule

$$UA_j = \bigcap_{k \geqslant j} N_{k,j} A_k \subset A_j.$$

The natural injections $i\colon A_j \to A_k$ preserve the submodules of universal norms, as do the mappings N.

[9] There is, perhaps, greater justification in referring to this as a *trace* rather than a norm, but our prototype is class field theory where A is the multiplicative group of a field.

If the A_j are $\boldsymbol{Z}_p$-modules of finite type, then the norm mappings

$$N\colon\ UA_k \to UA_j$$

are surjective. Let UA stand for the projective limit of the UA_k compiled via the norm mappings above, and viewed as a Λ-module.

18. Growth numbers. Using the theory of Λ-modules and [36], Prop. 6.4, one can show:

GROWTH NUMBER PROPOSITION. *Suppose that E has good ordinary reduction at every place v of K of characteristic p. Suppose further that for any v such that the Néron model of E at v is geometrically disconnected, v splits only finitely often in L.*

Then there is an integer a (independent of n) such that the L/K universal norm submodule $US_p(E, K_n) \subset S_p(E, K_n)$ is a free Λ_n-module of rank a, for all n.

We have the following asymptotic behavior for the ranks:

$$r_p(E, K_n) = a \cdot p^n + e_n \tag{*}$$

where the "error terms" e_n are nonnegative numbers monotonically increasing with n and admitting a uniform upper bound.

We refer to the integer $a = a(E, L/K)$ as the *growth* number of E for the $\boldsymbol{Z}_p$-extension L/K.

Let us return to $K = k$ a quadratic imaginary field and $E/\boldsymbol{Q}$ a Weil parametrized elliptic curve. Compatible with the conjectures and results quoted in Section 2, we have the

GROWTH NUMBER CONJECTURE. *Let p be a prime number of good, ordinary reduction for E. Suppose that the Néron fibre of E is geometrically connected at every place v of k at which the $\boldsymbol{Z}_p$-extension L/k splits infinitely. Then:*

$a(E, L/k) = 0$ *if L/k is not the anti-cyclotomic $\boldsymbol{Z}_p$-extension, or the sign of (E, χ_0) is $+1$. (Here χ_0 is the principal character over k.)*

If L/k is the anti-cyclotomic $\boldsymbol{Z}_p$-extension and the sign of (E, χ_0) is -1, then:

$$a(E, L/k) = 1 \quad \textit{in the generic case},$$

and:

$$a(E, L/k) = 2 \quad \textit{in the exceptional case}.$$

Remarks. (1) A more general growth number proposition could be proved concerning χ-eigenspaces for χ a tame character (i.e., of finite order prime to p) over k. We content ourselves here and in Sections 19, 20 and 22 below with statements for $\chi = \chi_0$, the principal character, although more general tame characters could also be treated.

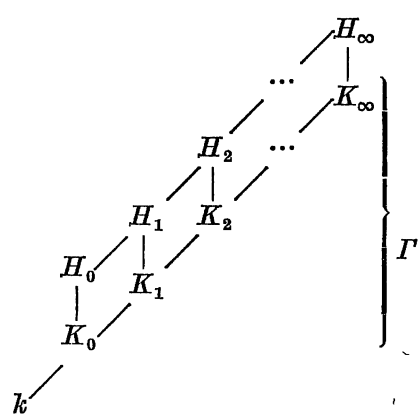
(2) See Monsky [41]–[43] for an analysis of possible growth behavior in $\boldsymbol{Z}_p[[\boldsymbol{Z}_p^d]]$ modules for general $d \geqslant 1$.

19. Heegner points as universal norms. Let $k \subset \boldsymbol{C}$ be a quadratic imaginary field, p a prime number, and N a positive integer prime to p, for which there exists an ideal $\mathscr{N}$ in the ring of integers $\mathcal{O}$ of k, of norm N. Fix such an ideal $\mathscr{N}$. Let $\mathcal{O}_n$ denote the order in $\mathcal{O}$ of conductor p^n, and $\mathscr{N}_n = \mathscr{N} \cap \mathcal{O}_n$. Let H_n be the ring class field of $\mathcal{O}_n$, so that $H_n = k(j(\mathcal{O}_n))$. If $H_\infty = \bigcup_n H_n$, then H_∞ is an anti-cyclotomic extension of k whose Galois group is isomorphic to $\boldsymbol{Z}_p \oplus F$ where F is a finite group; consequently, there is a unique $\boldsymbol{Z}_p$-extension K_∞/k contained in H_∞. The $\boldsymbol{Z}_p$-extension K_∞/k is the anti-cyclotomic $\boldsymbol{Z}_p$-extension of k.

Define

$$K_n = K_\infty \cap H_n \quad \text{for all } n \geqslant 0.$$

Thus K_0/k is a finite, cyclic, everywhere unramified extension of degree a power of p, and K_n/K_0 is a finite cyclic extension totally ramified at the primes dividing p, and of degree p^n.



Let $\varphi\colon X_0(N) \to E$ be a Weil parametrization of an elliptic curve $E/\boldsymbol{Q}$ (of conductor N).

Define the *submodule of Heegner points at the n-th layer*

$$\mathscr{E}(K_n) \subset [E(K_n) \otimes \boldsymbol{Z}_p]/\text{torsion}$$

to be the $\boldsymbol{Z}_p$-module generated by the trace to K_n of the basic Heegner points of level p^n. Thus, if

$$e_n^\sigma = \mathrm{Trace}_{H_n/K_n}\{x(\mathcal{O}_n, \mathcal{N}_n)^\sigma\}, \qquad \sigma \in \mathrm{Gal}(H_n/k),$$

then $\mathscr{E}(K_n)$ is the $\boldsymbol{Z}_p$-module generated by the e_n^σ $\big($for $\sigma \in \mathrm{Gal}(H_n/K)\big)$ in $E(K_n) \otimes \boldsymbol{Z}_p/\text{torsion}$.

Now let a_p denote the integer $1+p-\#\big(E(\boldsymbol{F}_p)\big)$, and let $\beta_p, \bar{\beta}_p$ denote the two roots of $X^2 - a_p X + p$. Suppose that p is *ordinary* for E, i.e., $a_p \not\equiv 0 \bmod p$, and that β_p is the p-adic unit root.

Let ε denote the quadratic character belonging to k, so that $1+\varepsilon(p)$ is the number of distinct primes of k with residue field $\boldsymbol{F}_p$. Let $M_p = \beta_p - 1 - \varepsilon(p)$. Then M_p is never zero.

Using formulas expressing the action of the Hecke operator T_p on Heegner points, an elementary calculation yields:

PROPOSITION. *Suppose that a_p is congruent to neither* 0 *nor* $1+\varepsilon(p)$ $\bmod p$. *Then working in the system of Γ-modules* $E(K_n) \otimes \boldsymbol{Z}_p/\text{torsion}$, *we have*

$$\mathrm{Norm}_{K_m/K_n} \mathscr{E}(K_m) = \mathscr{E}(K_n), \qquad m \geqslant n \geqslant 0.$$

More generally, if $a_p \not\equiv 0 \bmod p$, then $M_p \cdot \mathscr{E}(K_n)$ lies in the subspace of K_∞/K_n-universal norms.

For simplicity *we suppose, now, that $a_p \not\equiv 0$, or* $1+\varepsilon(p) \bmod p$. Define the *Heegner module*

$$\mathscr{E}(K_\infty) = \varprojlim_n \mathscr{E}(K_n)$$

which is the analogue for our elliptic curve E of the Λ-module constructed by taking the projective limit of the p-completion of the space of elliptic units in K_n, the limit being taken as n tends to ∞.

One easily sees that the Λ-module $\mathscr{E}(K_\infty)$ is cyclic, and since the natural mapping $\mathscr{E}(K_\infty) \to US_p(K_\infty)$ is nontrivial if $\mathscr{E}(K_\infty) \neq \{0\}$, and the latter Λ-module is free, it follows that $\mathscr{E}(K_\infty)$ is a free Λ-module of rank either 0 or 1. Clearly, if there exists some $n \geqslant 0$ for which e_n is nonzero, then $\mathscr{E}(K_\infty)$ is free of rank 1.

CONJECTURE. *There is some n for which e_n is nonzero; equivalently:* $\mathscr{E}(K_\infty)$ *is a free Λ-module of rank* 1.

Remarks. (1) The theory we have just described can be generalized with some modification to eigenspaces for some character χ of order prime to p. As suggested in Section 4, one should broaden the theory to include Heegner points coming from the various Shimura curve parametrizations of E as well.

(2) Kurcanov was the first to discover the possibility that Heegner points account for unbounded Mordell-Weil rank in anti-cyclotomic towers. See [32] for examples of nontriviality of Heegner modules attached to certain tame characters χ.

(3) It is worth noting that although the Λ-module structure of $\mathscr{E}(K_\infty)$ is simple enough, the structure of the finite layers $\mathscr{E}(K_n)$ can be quite subtle and, invoking standard conjectures, can be seen to reflect properties of the error term e_n in the formula

$$r_p(E, K_n) = a\cdot[K_n\colon k]+e_n.$$

20. p-adic heights. The classical height pairing on the Mordell–Weil group of an elliptic curve involves the choice of an "$\boldsymbol{R}$-valued logarithmic character"

$$\mathscr{A}_K^*/K^*\to\boldsymbol{R}^+$$

although the use of the "natural logarithm" obscures this choice.

A general "theory of p-adic heights" for elliptic curves E/K would yield a bilinear symmetric pairing

$$E(K)\times E(K)\to\boldsymbol{Q}_p,$$

$$(x, y)\to\langle x, y\rangle_\lambda$$

dependent upon a choice of p-adic valued logarithm λ over K (a *canonical λ-height pairing*) which is $\boldsymbol{Q}_p$-linear in λ, functorial, and, of course, satisfies the more elusive property of being an effective tool in the study of the arithmetic of E.

We do not yet have such a general theory, but much progress towards it has been recently made.

Say that a logarithm λ over K is *ordinary* for E if for every nonarchimedean prime v of K at which λ is ramified, either the Neron fibre of E is of multiplicative type or the "Trace of Frobenius" a_v is not congruent to zero modulo the characteristic of the residue field of v. Say that λ is *good* for E if for every v at which λ is ramified, E has good reduction.

Suggestions for a mod p theory first appeared in [53]. For elliptic curves of complex multiplication and for certain ordinary p-adic logarithms, Perrin-Riou [47] and Bernardi [3] developed an analytic theory. Perrin-Riou also produced an arithmetic (i.e., "descent-theoretic") definition in the same context and she established the equivalence of the two theories. More recently, Schneider [57] has given a general formulation of the theory of arithmetic height, valid for general elliptic curves E and good[10] ordinary logarithms for E. He also discovered another way of defining the height under the same hypothesis and has shown that his two definitions are equivalent. We refer to this as *Schneider's p-adic height*. Néron has yet another approach [44]. Gross and Oesterlé have a p-adic height for complex multiplication elliptic curves and good non-ordinary logarithms for E. In [38], a theory of p-adic heights for general elliptic curves and ordinary logarithms is developed; we refer to this as *analytic p-adic height.*

The Schneider λ-height pairing has the desirable property that its kernel contains the subspace of L/K-universal norms, where L/K is the $\boldsymbol{Z}_p$-extension cut out by the logarithm λ. The analytic height pairing, being "analytic", is amenable to numerical calculation. One can prove that Schneider's theory coincides with the analytic theory for good, ordinary, logarithms, but see the forthcoming [39] for a surprise in a more general context.

Now choose a basis $P_1, \ldots, P_r$ for the Mordell–Weil group $E(K)$ modulo torsion. Consider

$$\delta(\lambda) = \det_{ij}\langle P_i, P_j\rangle_\lambda,$$

viewed as giving rise to an r-linear symmetric form (*the height determinant form*) on the space of p-adic logarithms over k which are ordinary for E.

It would follow from the conjecture of Section 16 and the above discussion that if $\delta(\lambda) \neq 0$, i.e., the λ-height pairing is nondegenerate, then the growth number of L/K, the $\boldsymbol{Z}_p$-extension cut out by λ, vanishes.

What r-linear forms can arise as height determinant forms? See forthcoming work of G. Brattström related to this question.

21. The (p-adic valued) analytic theory. If χ is a Hecke character of finite order over k, let $\boldsymbol{Q}(\chi) \subset \boldsymbol{C}$ denote the subfield generated by the values of χ. By the recent work of Cogdell–Stevens [9], there is a nonzero complex number Ω_E (the period) such that the special values

$$\Lambda(E, \chi) \underset{\text{defn.}}{=} L(E, \chi, 1)/\Omega_E$$

lie in the field $\boldsymbol{Q}(\chi)$.

[10] Schneider's theory also covers certain cases of bad ordinary reduction.

Now let Δ be any integer, p a prime of good, ordinary reduction for E and $G_{p,\Delta}$ the Galois group of the union of all abelian extensions of k of conductor dividing $p^m\cdot\Delta$ for some m. From the work of Kurcanov, following that of Manin (See also Haran [24]), one sees that there is a Q_p-valued measure $\mu_{p,\Delta}$ on $G_{p,\Delta}$ (i.e., a distribution with bounded values) which "p-adically interpolates" the special values $\Lambda(E,\chi)$ in the sense that we have a formula of the form:

$$\int_{G_{p,\Delta}} \chi\cdot\mu_{p,\Delta} = c_{p,\Delta}(E,\chi)\cdot\Lambda(E,\chi)$$

valid in $Q(\chi)\otimes Q_p$, where χ is any character of k of conductor dividing $p^m\Delta$ for some m (viewed on the left-hand side as a character on $G_{p,\Delta}$), and where $c_{p,\Delta}(E,\chi)$ is a nonzero "elementary term"; elementary in the sense that its dependence on E is purely local: it can be expressed in terms of the Euler factors L_v of E for v dividing $p\Delta$.

For a fixed character χ of conductor equal to Δ as above, and a p-adic logarithm λ, we may define the p-adic L function (in the λ direction) as the power series in $Q(\chi)\otimes Q_p[[X]]$:

$$L_\lambda(\chi;X) = \sum_{n=0}^{\infty}\frac{1}{n!}\int_{G_{p,\Delta}}\lambda^n\cdot\chi\cdot\mu_{p,\Delta}\cdot X^n$$

and, making use of our choice of cyclotomic and anti-cyclotomic p-adic logarithms σ and τ, we define the *two-variable p-adic L function* by:

$$L_{\sigma,\tau}(\chi;S,T) = \sum_{n,m\geqslant 0}\frac{1}{n!\cdot m!}\int_{G_{p,\Delta}}\chi\cdot\sigma^n\cdot\tau^m\cdot\mu_{p,\Delta}\cdot S^nT^m.$$

Clearly, the substitution $S=aX$, $T=bX$ in the two-variable p-adic L series yields the p-adic L series in the $\lambda=a\sigma+b\tau$ direction.

The results on p-adic interpolation are obtained by working with modular symbols attached to lifts of the automorphic form determined by the Weil parametrized elliptic curve E.

There has been an impressive output of work devoted to the establishment of a *one-variable* theory of p-adic L-functions (principally in the cyclotomic direction but see [6], [7] for very interesting results in different directions). As for the two-variable theory, prior to the work of Kurcanov and Haran, Katz [29], [30] and Manin–Visik [35] defined a p-adic L function attached to Grossencharacters over a quadratic imaginary field, by interpolation of special values of certain Eisenstein series and Yager [66], [68], [69] provided an alternative approach to Katz's L-functions by considering the images of elliptic units in local units.

If $\vee$ denotes k/Q-conjugation, since one can show that $\mu_{p,\Delta}^{\vee} = \mu_{p,\Delta}$, we have:

$$L_{\sigma,\tau}(\chi; S, T) = L_{\sigma,\tau}(\chi; S, -T).$$

We also have the functional equation for characters χ of conductor $\mathfrak{f}$ prime to N which can be written

$$L_{\sigma,\tau}(\chi; S, T) = e(E, \chi)\cdot\exp\big(S\cdot\sigma(N)\big)\cdot\chi(N)\cdot L_{\sigma,\tau}(\chi^{-1}; -S, -T)$$

for a suitable constant $e(E, \chi)$. Combining the above two equations *when χ is anti-cyclotomic* gives:

$$L_{\sigma,\tau}(\chi; S, T) = e(E, \chi)\cdot\exp\big(S\cdot\sigma(N)\big)\cdot L_{\sigma,\tau}(\chi; -S, T)$$

again under the hypothesis that $\mathfrak{f}$ is prime to N. The constant $e(E, \chi)$ is the sign of (E, χ) in the generic case.

From the functional equation one readily sees that when the sign of (E, χ) is -1 (χ anti-cyclotomic, $\mathfrak{f}$ prime to N), the anti-cyclotomic L series $L_\tau(\chi; X)$ vanishes identically. In this case, define the *first derived anti-cyclotomic* L series by the formula:

$$L_\tau'(\chi; X) \underset{\text{defn.}}{=} d/dS\big(L_{\sigma,\tau}(S, X)\big)\big|_{S=0}.$$

In the exceptional case, when the sign is -1, the first derived anti-cyclotomic L series will also vanish and it is the second derived L series which will be of interest.

Concerning the two-variable p-adic L function, it would be very interesting to frame precise conjectures connecting it to the arithmetic of E along the lines that have been explored in analogy with conjectures already formulated in the one-variable theory.

Notably, one might hope for a "main conjecture" which relates the locus of zeros of $L_{\sigma,\tau}$ in a suitable ball to the divisorial part of the support of an Iwasawa module built from p-primary Selmer groups. One might hope for a "Birch–Swinnerton-Dyer-type conjecture" which would assert that if the two-variable p-adic L series is written as an infinite sum

$$\sum_{j\geqslant\nu} l_{\sigma,\tau}^{(j)}(\chi; S, T)$$

where $l_{\sigma,\tau}^{(j)}(\chi; S, T)$ is a homogeneous polynomial of degree j in the variables S and T and $l_{\sigma,\tau}^{(\nu)}(\chi; S, T)$ is the lowest nonvanishing term, then $\nu = r(E, \chi)$ and $l^{(\nu)}$ is a scalar multiple of a suitable "height determinant form" (cf.

Section 20). One might hope to pin down that scalar multiple. One would expect a "theory of congruences modulo Eisenstein primes" along the lines of the one-variable theory [37], [61].

22. p-adic valued Heegner measures. Let us return to the setting of Section 19. Choose a generator c_∞ of the Heegner module $\mathscr{E}(K_\infty)$. Let $c_n \in \mathscr{E}(K_n)$ be the image of c_∞. Define a Q_p-valued measure ν_n on Γ_n by the rule:

$$\nu_n(g) = p^n \langle c_n, c_n^g \rangle_\sigma,$$

where $\langle , \rangle_\sigma$ denotes the p-adic valued height attached to σ, the p-cyclotomic logarithm.

One easily checks that the $(\nu_n)_n$ satisfy the distribution law

$$\nu_n(g) = \sum_{g'} \nu_{n+1}(g')$$

where the sumation runs through all elements of Γ_{n+1} projecting to g in Γ_n. Thus the $(\nu_n)_n$ determine a distribution ν on Γ, which can be seen (using the results of [38]), to be a measure, i.e., $\nu^{(c_\infty)} \in \Lambda \otimes_{Z_p} Q_p$. Changing the choice of generator c_∞ of (K_∞) changes $\nu^{(c_\infty)}$ by multiplication by a unit in Λ.

Is there a p-adic version of the theory of Gross and Zagier, which relates the first derived p-anti-cyclotomic L-function $L'_\tau(\chi_0, X)$ to the Heegner measure $\nu^{(c_\infty)}$?

References

[1] Asai T., On Certain Dirichlet Series Associated With Hilbert Modular Forms and Rankin's Method, *Math. Ann.* **226** (1977), pp. 81–94.

[2] Asai T., On the Doi–Naganuma Lifting Associated With Imaginary Quadratic Fields, *Nagoya Math. J.* **71** (1978), pp. 149–167.

[3] Bernardi D., Hauteur p-adique sur les courbes elliptiques. Séminaire de Théorie des Nombres, Paris 1979-80, Séminaire Delange-Pisot-Poitou. *Progress in Math* **12**, Birkhäuser, Boston–Basel–Stuttgart (1981), pp. 1–14.

[4] Birch B., Elliptic Curves and Modular Functions, *Symposia Mathematica, Inst. di Alta Mathematica*, **IV** (1970), pp. 27–32.

[5] Birch B., Heegner Points of Elliptic Curves, *Symposia Mathematica, Inst. di Alta Mathematica,* **XV** (1975), pp. 441–445.

[6] Coates J., Infinite Descent on Elliptic Curves in Arithmetic and Geometry. Papers dedicated to I. R. Shafarevich on the occasion of his sixtieth birthday. Volume I. *Progress in Math.* **35**, Birkhäuser, Boston–Basel–Stuttgart (1983), pp. 107–136.

[7] Coates J., *Arithmetic on Elliptic Curves With Complex Multiplication,* Hermann Weyl Lectures, I. A. S., Princeton 1979, to appear.

[8] Coates J. and Wiles A., On the Conjecture of Birch and Swinnerton-Dyer, *Inv. Math.* **39** (1977), pp. 223–251.

[9] Cogdell J. and Stevens G., to appear.

[10] Deligne P., Valeurs de fonctions L et périodes d'intégrales. In: *Automorphic Forms, Representations, and L-functions. Proceedings of the symposium in pure mathematics of the A.M.S.* (Oregon State U., Corvallis, Oregon 1977), Vol. XXXIII, Part 2, A.M.S. Providence, 1979, pp. 313–346.

[11] Dirichlet L., Sur la manière de résoudre l'equation $t^2 - pu^2 = 1$ au moyen des fonctions circulaires, *Werke*, Vol. I, Georg Reimer, Berlin, 1889, pp. 345–350.

[12] Friedman E., Ideal Class Groups in Basic $\boldsymbol{Z}_{p_1} \times \ldots \times \boldsymbol{Z}_{p_s}$-Extension of Abelian Number Fields, *Inv. Math.* **65** (1982), pp. 425–440.

[13] Goldfeld D., The Class Number of Quadratic Fields and the Conjectures of Birch and Swinnerton-Dyer, *Annali Scuola Normale Superiore – Pisa*, Serie IV, Vol. III, 4 (1976), pp. 623–663.

[14] Goldfeld D., Conjectures on Elliptic Curves Over Quadratic Fields. In: *The Proceedings of the Conference on Number Theory, Carbondale, 1979*, Lecture Notes in Math. **751**, Springer, Berlin–Heidelberg–New York (1979), pp. 108–118.

[15] Goldfeld D. and Viola C., Mean Values of L-Functions Associated to Elliptic, Fermat and Other Curves at the Center of the Critical Strip, *Journal of Number Theory* **11** (1979), pp. 305–320.

[16] Goldfeld D., Hoffstein J. and Patterson S., On Automorphic Functions of Half-Integral Weight With Applications to Elliptic Curves, in Number Theory Related to Fermat's Last Theorem. Proceedings of the Conference sponsored by the Vaughn Foundation, *Progress in Math.* **26**, Birkhäuser, Boston–Basel–Stuttgart (1982), pp. 153–193.

[17] Greenberg R., On the Birch and Swinnerton-Dyer Conjecture, *Inv. Math.* **72** (1983), pp. 241–265.

[18] Greenberg R., *On the Critical Values of Hecke L-Functions for Imaginary Quadratic Fields*, preprint, University of Washington.

[19] Greenberg R., Iwasawa's Theory and p-Adic L-Functions for Imaginary Quadratic Fields. Number Theory related to Fermat's Last Theorem. Proceedings of the Conference sponsored by the Vaughn Foundation, *Progress in Math.* **26**, Birkhäuser, Boston–Basel–Stuttgart (1982), pp. 275–286.

[20] Gross B., *Heegner Points on $X_0(N)$*, preprint, Brown University.

[21] Gupta R., *Fields of Division Points of Elliptic Curves Related to Coates-Wiles*, M.I.T., Ph.D. Thesis, 1983.

[22] Gross B. and Zagier D., Points de Heegner et dérivées de fonctions L, *Comptes Rendus Acad. Sc. Paris* **297** (1983), pp. 85–87.

[23] Gross B. and Zagier D., further publications.

[24] Haran S., *p-Adic L-Functions for Elliptic Curves Over CM fields*, M.I.T., Ph.D. Thesis, 1983.

[25] Heegner K., Diophantische Analysis und Modulfunktionen, *Math. Zeitschrift* **56** (1952), pp. 227–253.

[26] Jacquet H., *Automorphic Forms on* GL (2), II, Springer Lecture Notes **278** (1972).

[27] Jacquet H. and Langlands R., *Automorphic Forms on* GL (2), Lecture Notes in Math. **114**, Springer, Berlin–Heidelberg–New York (1970).

[28] Jordan B., *On the Diophantine Arithmetic of Schimura Curves*, Harvard, Ph.D. Thesis, 1981.

[29] Katz N., p-Adic Interpolation of Real Analytic Eisenstein Series, *Ann. of Math.* **104** (1976), pp. 459–571.
[30] Katz N., p-Adic L-Functions for CM fields, *Inv. Math.* **49** (1978), pp. 199–297.
[31] Kurcanov P., Elliptic Curves of Infinite Rank Over Γ-Extensions, *Mat. Sbornik* **93** (135) (1974), pp. 460–466.
[32] Kurcanov P., The Zeta-Function of an Elliptic Curves Over Certain Abelian Extensions of Imaginary Quadratic Fields, *Mat. Sbornik* **31** (1977), pp. 49–62.
[33] Kramer K. and Tunnell J., Elliptic Curves and Local ε-Factors, *Comp. Math.* **46** (1982), pp. 307–352.
[34] Lang S., *Elliptic Curves and Diophantine Analysis*, Springer, 1978.
[35] Manin J. and Visik H., p-Adic Hecke Series for Quadratic Imaginary Fields, *Mat. Sbornik* **95** (1974), pp. 357–383.
[36] Mazur B., Rational Points of Abelian Varieties With Values in Towers of Number Fields, *Inv. Math.* **18** (1972), pp. 183–266.
[37] Mazur B., On the Arithmetic of Special Values of L-Functions, *Inv. Math.* **55** (1979), pp. 207–240.
[38] Mazur B. and Tate J., Canonical Height Pairings via Biextensions, in Arithmetic and Geometry. Papers dedicated to I. R. Shafarevitch on the occasion of his sixtieth birthday, Volume I. *Progress in Math.* **35**, Birkhäuser, Boston–Basel–Stuttgart (1983), pp. 195–237.
[39] Mazur B., Tate J. and Teitelbaum J., *On the p-Adic Birch Swinnerton-Dyer Conjecture*, in preparation.
[40] *Modular Functions of one variable* IV. Proceedings of the International Summer School, University of Antwerp, RUCA (1972), Lecture Notes in Math. **467**, Springer, Berlin–Heidelberg–New York (1975).
[41] Monsky P., *The Hilbert–Kunz Function*, preprint.
[42] Monsky P., *p-Ranks of Class Groups in* $\mathbf{Z}_p^d$*-Extensions*, preprint.
[43] Monsky P., *High Growth and Ordinary Growth of Class Numbers in* $\mathbf{Z}_p^d$*-Extensions*, preprint.
[44] Néron A., Hauteurs et fonctions thêta, *Rend. Sci. Mat. Milano* **46** (1976), pp. 111–135.
[45] Oesterlé J., Construction de hauteurs archimédiennes et p-adiques suivant la méthode de Bloch. Séminaire de Théorie des Nombres (Séminaire Delange-Pisot-Poitou), Paris, 1980-81. *Progress in Math.* **22**. Birkhäuser, Boston–Basel–Stuttgart (1982), pp. 175–192.
[46] Perrin-Riou B., Groupe de Selmer d'une courbe elliptique à multiplication complexe, *Comp. Math.* **18** (1972), pp. 183–266.
[47] Perrin-Riou B., Descente infinie et hauteur p-adique sur les courbes elliptiques à multiplication complexe, *Inv. Math.* **70** (1983), pp. 369–398.
[48] Ramachandra K., Some Applications of Kronecker's Limit Formulas, *Ann. of Math.* **80** (1964), pp. 104–148.
[49] Robert G., Unitées elliptiques, *Bull. Soc. Math. France. Mémoire* **36** (1973).
[50] Rohrlich D., *On L-Functions of Elliptic Curves and Anticyclotomic Towers*, preprint.
[51] Rohrlich D., The Nonvanishing of Certain Hecke L-Functions at the Center of the Critical Strip, *Duke Math. J.* **47** (1980), pp. 223–232.
[52] Rohrlich D., On the L-Functions of Canonical Hecke Characters of Imaginary Quadratic Fields, *Duke Math. J.* **47** (1980), pp. 547–557.

[53] Rosenblum M. in collaboration with Abramov S., *The Birch Swinnerton-Dyer Conjecture* mod p, handwritten manuscript.
[54] Rubin K., Congruences for Special Values of L- Functions of Elliptic Curves with Complex Multiplication, *Inv. Math.* **71** (1983), pp. 339–364.
[55] Rubin K., Congruences for Special Values of L- Functions of Elliptic Curves With Complex Multiplication, *Inv. Math.* **71** (1983), pp. 339–364.
[56] Rubin K. and Wiles A., Mordell–Weil Groups of Elliptic Curves Over Cyclotomic Fields. Number theory related to Fermat's Last Theorem, *Progress in Math.* **26**, Birkhäuser, Boston–Basel–Stuttgart (1982), pp. 237–254.
[57] Schneider P., Iwasawa L-Functions of Varieties Over Algebraic Number Fields. A First Approach, *Inv. Math.* **71** (1983), pp. 251–293.
[58] Schneider P., p-Adic Height Pairings I, *Inv. Math.* **69** (1982), pp. 401–409.
[59] Stark H., The Coates–Wiles Theorem Revisited. Number theory related to Fermat's Last Theorem, *Progress in Math.* **26**, Birkhäuser, Boston–Basel–Stuttgart (1982), pp. 349–362.
[60] Stark H., L-Functions at $s = 1$. IV. First Derivatives at $s = 0$. *Advances in Math.* **35** (1980), pp. 197–235.
[61] Stevens G., Arithmetic on Modular Curves, *Progress in Math.* **20**, Birkhäuser, Boston–Basel–Stuttgart, 1982.
[62] Vigneras M.-F., Valeur au centre de symétrie des fonctions L associées aux formes modulaires in Séminaire de Théorie des Nombres, Paris 1979–80. Séminaire Delange–Pisot–Poitou. Birkhäuser, Boston–Basel–Stuttgart. *Progress in Math.* **12** (1981), pp. 331–356.
[63] Waldspurger J.-L., *Correspondence de Shimura et quaternions*, preprint.
[64] Washington L., The Non-p-Part of the Class Numbers in a Cyclotomic $\boldsymbol{Z}_p$-extension, *Inv. Math.* **49** (1978), pp. 87–97.
[65] Weil A., *Dirichlet Series and Automorphic Forms*, Lecture Notes in Math. **189**, Springer, Berlin–Heidelberg–New York (1971).
[66] Yager R., On Two-Variable p-Adic L-Functions, *Ann. of Math* **115** (1982), pp. 411–449.
[67] Yager R., A Kummer Criterion For Imaginary Quadratic Fields, *Comp. Math.* **47** (1982), pp. 31–42.
[68] Yager R., *p-Adic Measures on Galois Groups*, preprint, M.I.T.
[69] Yager R., *Iwasawa Theory For the Anticyclotomic Extension*, preprint, M.I.T.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

ROBERT MACPHERSON

Global Questions in the Topology of Singular Spaces

During the last decade, several new tools have been found for the study of the topology of singular spaces. Our object here is to review the part of this work that relates to intersection homology.

When he introduced homology for the study of manifolds, Poincaré made it clear that he was motivated by applications in three directions: analysis (differential equations), algebraic geometry, and group theory ([68], p. 194). Each of these three fields leads to questions about singular spaces as well. After developing some general theory, we will illustrate its usefulness in three sections devoted to applications in these three areas.

The singular spaces that arise in applications are usually complex algebraic (or analytic) varieties. In this report, we will restrict ourselves to methods which apply to these. Even here, desire for unity, limitations of space, and ignorance force the omission of many important subjects. For example, we do not mention the rather well developed theory of characteristic classes and Riemann–Roch (see [73], [59], [9], [10], [30], [31], [32]).

§ 1. Stratifications. Perhaps the primary reason why the study of singular spaces blossomed at this time was the creation of stratification theory [58], [80], [74], [62]. This illuminated the local structure of analytic varieties.

Let X be a complex analytic or algebraic variety of pure complex dimension n. Then X admits a locally finite decomposition into disjoint connected nonsingular analytic subvarieties $\{S_a\}$ of varying dimension called *strata*, which satisfies a *homogeneity condition* along the strata: for any two points p and q on a stratum S_a, there exists a homeomorphism of X to X, preserving all the strata and taking p to q. We denote the codimension of S_a by c_a, i.e. $c_a = n - \dim S_a$. The space Σ which is the union of all S_a such that $c_a > 0$ contains all the singularities of X. The strata $S_a \subset \Sigma$ may be thought of as loci on which X is "uniformly singular".

This homogeneity of X along S_α is guaranteed by the Whitney conditions: (1) the closure $\bar{S}_\alpha$ of S_α is a union of strata and (2) if a sequence of points $b_i \in S_\beta$ and a sequence $a_i \in S_\alpha$ both approach the same point $c \in S_\beta$, then the limit of the secant lines connecting a_i to b_i is contained in the limit of the tangent spaces at a_i, if both limits exist.

Intersection homology

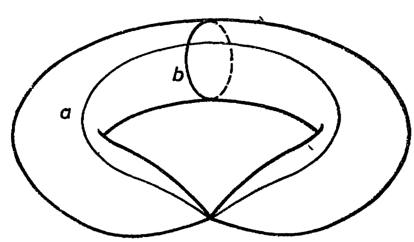
§ 2. Motivation. One of the important properties of the ordinary homology $H_i(M)$ of a compact oriented $2n$ real dimensional manifold M is *Poincaré duality*:

(1) *If $i+j = 2n$, then any pair of homology classes $A \in H_i(M)$ and $B \in H_j(M)$ have representative cycles $a \in A$ and $b \in B$ that intersect in finitely many points.*

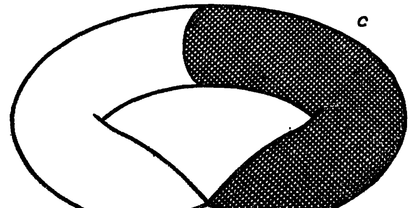
(2) *The number of these intersection points counted according to their multiplicities is independent of the choice of a and b and is denoted $A \cap B$.*

(3) *If M is compact, the bilinear pairing $H_i(M, Q) \times H_j(M, Q) \to Q$ which sends (A, B) to $A \cap B$ is nondegenerate.* (See [53] for the definition of intersection multiplicities.)

An obvious extension of $H_i(M)$ to singular varieties X is the usual homology group $H_i(X)$. Part (2) of the Poincaré duality fails even for the simplest singular space; the nodal cubic.



From the position of a and b in the picture, we would conclude that $A \cap B = 1$. However, b is the boundary of a chain c so $B = 0$; hence $A \cap B = 0$.



Another obvious extension of $H_i(M)$ to singular X would be $H^{2n-i}(X)$ which coincidens with it in the nonsingular case. By viewing $H^{2n-i}(X)$ as traces in X of bigger cycles in an ambient nonsingular space, one may visualize $H^{2n-i}(X)$ as the homology of a complex of chains which satisfy a transversality condition with respect to the singularities of X (see [37]). If X is the nodal cubic, a does not satisfy this condition, and neither does the chain c. So $H^1(X)$ is one-dimensional, generated by B. Property (2) of Poincaré duality holds but (3) fails since no class pairs nontrivially to B.

In the definition of intersection homology, we will restore Poincaré duality to the singular case by placing conditions on how chains meet the singular strata which are less strict than the transversality condition for $H^{2n-i}(X)$. For the nodal cubic, a will not be allowed but c will, so the first intersection homology group is zero.

§ 3. Definition of intersection homology. Let X be an n-dimensional complex analytic variety with a Whitney stratification $\{S_a\}$, and let c_a be the complex codimension of S_a. Denote by $\{C_i(X)\}$, $i \in \mathbf{Z}$, the complex of geometric chains on X. (Several choices for this work equally well. For example, we can take $C_i(X)$ to be the piecewise linear i-chains with respect to some piecewise linear structure on X. An element c of $C_i(X)$ would then be a simplicial chain for some triangulation (depending on c) of X. Another choice would be subanalytic chains [45].)

We define the complex of intersection chains $\{IC_i(X)\}$ to be the subcomplex of $\{C_i(X)\}$ consisting of those chains $c \in C_i(X)$ satisfying the *allowability condition* ([39], [3]):

The chain c intersects each singular stratum $S_a \subset \Sigma$ in a set of (real) dimension less than $i - c_a$ and its boundary ∂c intersects each singular stratum $S_a \subset \Sigma$ in a set of dimension less than $i-1-c_a$. (Note that c_a, the complex codimension of S_a, is half the real codimension.)

For some purposes dimension bounds other than $i-c_a$ and $i-1-c_a$ are useful, and the chains described above are sometimes called middle-perversity intersection chains in the literature. We will not use these other dimension bounds here.

DEFINITION [39]. The ith *intersection homology* group of X, denoted $IH_i(X)$, is the ith homology group of the chain complex $\{IC_i(X)\}$.

If X is nonsingular, then $IH_i(X) = H_i(X)$. In general the groups $IH_i(X)$ have many attributes of more familiar topological functors. They depend only on the topology of S — in particular they are independent of the stratification $\{S_a\}$ used in their definition [4]. If X is compact, they are finitely generated. If $U \subset X$ is open, there are relative groups $IH_i(X, U)$ which fit in the usual long exact sequence and satisfy excision. Over a field F, they satisfy the Kunneth theorem

$$IH_i(X \times Y; F) = \bigoplus_{j+k=i} IH_j(X; F) \otimes IH_k(Y, F).$$

They are not homotopy invariants, but they are both covariant and contravariant functors on a class of maps called placid maps.

DEFINITION. A map of a purely m-dimensional variety Y^m to X^n is *placid* if X can be stratified by strata $\{S_a\}$ so that for each a, $\dim_C f^{-1}(S_a) \leqslant m - c_a$.

If f is placid, the homomorphisms

$$f_*: IH_i(Y) \to IH_i(X) \quad \text{and} \quad f^*: IH_i(X) \to IH_{i+2(m-n)}(Y)$$

are defined as usual, essentially by image cycles and transverse inverse image cycles.

§ 4. The Kähler package. The intersection homology $IH_i(X)$ of a singular algebraic variety X satisfies a large part of the package of special properties of the ordinary homology of a Kähler manifold. These results are all false for $H_i(X)$.

POINCARÉ DUALITY [3]. (1) *If $i+j = 2n$, then any pair of intersection homology classes $A \in IH_i(X)$ and $B \in IH_j(X)$ have representatives $a \in A$ and $b \in B$ that intersect only in $X - \Sigma_n$ and in finitely many points.*

(2) *The number of these intersection points counted according to their multiplicities is independent of the choice of a and b, and is denoted $A \cap B$.*

(3) *If X is compact, the bilinear pairing*

$$IH_i(X; Q) \times IH_j(X; Q) \to Q$$

which sends A, B to $A \cap B$ in nondegenerate.

The corresponding pairing over the integers is not unimodular [38]. For general i and j, there is an intersection pairing $IH_i(X) \times IH_j(X) \to H_{i+j-2n}(X)$ but no ring structure on $IH_*(X)$.

LEFSCHETZ HYPERPLANE THEOREM [4], [41]. *Let X^n be a closed subvariety of complex projective m-space CP^m and let $H^{m-1} \subset CP^m$ be a generic hyperplane. Then*

$$j_*: IH_i(X \cap H; Z) \to IH_i(X; Z)$$

is an isomorphism for $i < n-1$ *and is surjective for* $i = n-1$, *where* $j: X \cap H \hookrightarrow X$ *is the (placid) inclusion.*

HARD LEFSHETZ THEOREM [27], [12]. *Let* X *be a closed subvariety of* CP^m. *Then intersecting with a generic hyperplane* $H \subset CP^m$ *induces a mapping* $\bigcap [H]: IH_i(X) \to IH_{i-2}(X)$, *and for all* k *the iterated map*

$$IH_{n+k}(X; Q) \xrightarrow{(\cap[H])^k} IH_{n-k}(X; Q)$$

is an isomorphism.

CONJECTURE [26]. *If* X *is compact and projective, then* $IH_k(X; C)$ *has a pure Hodge decomposition*

$$IH_k(X; C) = \bigoplus_{i+j=k} IH_{i,j}(X)$$

with the following properties:

(a) $IH_{i,j}(X) = \overline{IH_{j,i}(X)}$;

(b) *If* $f: Y^m \to X^n$ *is placid,* $f_* IH_{i,j}(Y) \subset IH_{i,j}(X)$ *and* $f^* IH_{i,j}(X) \subset IH_{i+m-n, j+m-n}(Y)$;

(c) *The usual relations with the Lefschetz map* $\bigcap [H]$ *and the duality pairing* $A \cap B$ *hold; in particular the Hodge index theorem is valid.*

§ 5. Interpretation of intersection homology. Since the thrust of the last section (and the next two) is that $IH_*(X)$ behaves in many circumstances exactly like the ordinary homology of a nonsingular variety, one might well ask if $IH_*(X)$ is in fact the homology of an associated non-singular variety. The answer is: Sometimes.

A *small* resolution $\tilde{X} \to X$ is a resolution for which X can be stratified by strata $\{S_a\}$ such that if $p \in S_a$, then $\dim f^{-1}(p) < \frac{1}{2} c_a$, where c_a is the codimension of S_a in X. If $\pi: \tilde{X} \to X$ is a small resolution, then π induces an isomorphism $H_i(\tilde{X}) \cong IH_i(X)$ (see [40]).

This observation leads to the following fanciful question: In such a situation, how much of the topology of $\tilde{X}$ can be read from that of X? More precisely, what invariants I can be defined for all singular spaces so that, whenever a small resolution $\tilde{X}$ exists, $I(X) = I(\tilde{X})$? The list includes $H_*(\tilde{X})$ with its intersection bilinear form, the Wu class in $H_*(\tilde{X})$ [38], and the Chern numbers of $\tilde{X}$ for the signature, the Euler characteristic, and the arithmetic genus. But a bound on such speculation is provided by the fact that X may have two different small resolutions $\tilde{X}_1$ and $\tilde{X}_2$. Examples exist where the intersection ring structures of $\tilde{X}_1$ and

$\tilde{X}_2$ are not isomorphic and the Chern classes are not correspondingly placed. (This shows that there can be no natural ring structure on intersection homology.)

§ 6. Stratified Morse theory. We present here what appears to be the correct analogue for singular analytic varieties of Morse theory for manifolds [41]. Suppose X is embedded in a smooth complex analytic variety M and is Whitney stratified by $\{S_a\}$. For each stratum S_a, define the *conormal space* $C(S_a)$ to be the closure in T^*M, the cotangent bundle of M, of the set of cotangent vectors which lie over S_a and annihilate all tangent vectors in TS_a, the tangent space to S_a. An *X-critical point* of a smooth function $f\colon M\to \boldsymbol{R}$ is a critical point $p\in S_a$ of the restriction of f to some stratum S_a. The *critical value* of f at a critical point p is $f(p)$. A smooth function $f\colon M\to\boldsymbol{R}$ is called *Morse* for X if

(a) The restriction f/S_a is Morse for all $S_a\subset X$.

(b) If $p\in S_a$ is an X-critical point, $df(p)\notin C(S_\beta)$ for any $\beta\neq a$.

(c) The critical values of f are distinct.

The set of Morse functions for X is open and dense in the C^∞ topology; Morse functions are C^0 structurally stable on X [67]. For any number $s\in\boldsymbol{R}$ we denote by $X_{<s}$ the set of $x\in X$ such that $f(x)<s$. If $p\in S_a$ is a critical point for the Morse function f, we define the Morse index λ_p for f at p to be c_a+ (the Morse index for $f|S_a$ at p).

THEOREM. *There exists a unique set of abelian groups A_a one for each stratum S_a, such that for any proper Morse function $f\colon M\to\boldsymbol{R}$*

(a) *if the interval $[s,t)$ contains no critical values, then $IH_i(X_{<t}, X_{<s}; \boldsymbol{Z})=0$ for all i;*

(b) *if the interval $[s,t)$ contains the critical value v of one critical point $p\in S_a$, and λ_p is the Morse index of f at p, then*

$$IH_i(X_{<t}, X_{<s}; \boldsymbol{Z}) = \begin{cases} 0 & \text{for } i\neq\lambda_p,\\ A_a & \text{for } i=\lambda_p.\end{cases}$$

There is no analogous notion of a Morse index for ordinary homology replacing intersection homology since $H_i(X_{<t}, X_{<s})$ may be nonzero for several i, even if $[s,t)$ contains only one critical value.

The groups A_a are very difficult to calculate, but they are important since they arise in a number of other contexts. If m_a denotes the rank of the free part of A_a, the algebraic cycle in T^*M

$$\mathrm{ch}(X)=\sum_a m_a C(S_a)$$

is called the *characteristic variety* of X. It will play a role in § 7 and § 11 below.

§ 7. Lefschetz fixed point theorem. If $f: X \to X$ is a placid self-map, for example a self-homeomorphism, then the intersection homology Lefschetz number $IL(f) = \sum(-1)^i \operatorname{trace}\big(f_*: IH_i(X, Q) \to IH_i(X; Q)\big)$ has an expression which is localized at the fixed point set of f. More precisely, for each connected component K of the fixed point set there is a Lefschetz index $IL(f, K)$ determined by the local behavior of f near K, such that the sum of the $IL(f, K)$ over all connected components K is $IL(f)$. We give two formulas for $IL(f, K)$, both analogues of classical formulas for manifolds.

First we treat continuous placid self-maps $f: X \to X$ and give a result in the framework of [76]. Given any non-singular point $p \in X$, the cycle $[p \times X] \in X \times X$ satisfies the allowability conditions of § 2 and therefore lies in the intersection homology group $IH_{2n}(X \times X)$.

THEOREM [42]. *Let U_Δ and U_f be open regular neighborhoods of the diagonal and the graph of f respectively in $X \times X$. Then there are unique intersection homology classes with compact support $[\Delta] \in IH_{2n}(U_\Delta; Q)$ and $[f] \in IH_{2n}(U_f; Q)$ such that $[\Delta] \cap [p \times X] = 1$* and *$[f] \cap [p \times X] = 1$ for all non-singular $p \in X$. For these classes, we have*

$$IL(f) = [\Delta] \cap [f].$$

We may choose U_Δ and U_f so small that $U_\Delta \cap U_f$ has a unique connected component containing each connected component $\Delta \cap K \times K$ of $\Delta \cap X$ graph f. In this way $[\Delta] \cap [f]$ may be considered as a sum of contributions from each K, and we obtain a formula for $IL(f, K)$ as well. For example, if p is an isolated fixed point of f, then $IL(f, p)$ is the linking number of $L \cap [\Delta]$ and $L \cap [f]$ in L = the link of $p \times p$ in $X \times X$. (One can verify that linking numbers exist for disjoint $(2n-1)$-dimensional intersection homology cycles in L.)

Second we treat integrals of vector fields and give a generalization of the Hopf index formula. Suppose X is embedded in a smooth m-dimensional complex variety M, and let $v: X \to TM$ be a vector field on X, i.e. a (possibly discontinuous) section of TM defined only over X, such that if $p \in S_\alpha$, $v(p) \in TS_\alpha$. We suppose that v can be integrated (i.e. there is a one-parameter family $f_t: X \to X$ of self-maps of X such that f_0 is the identity and $\partial/\partial t\, f_t(p) = v\big(f_t(p)\big)$) and that the fixed points of f_t are exactly the zeros of v for $t \in (0, 1]$. "Controlled vector fields" (see [62]) provide a rich supply of such v. Choose a continuous section s: $M \to T^*M$

of the cotangent bundle to M with the property that if $p \in X$ and $v(p) \neq 0$ then $s(p)v(p) > 0$.

The image of s is a $2m$-cycle with closed supports $[s(M)]$ in T^*M. Another natural $2m$-cycle in T^*M is $\mathrm{ch}(X)$, defined at the end of § 6. Now given a connected component K of the fixed point set of f_1, pick an open set $U \subset M$ containing K but no other fixed points. The cycles $[s(M)]$ and $\mathrm{ch}(X)$ restrict to cycles $[s(U)]$ and $\mathrm{ch}(X, U)$ with closed support in $T^*(U)$. One can check that the condition $s(p)v(p) > 0$ for $p \in U-K$ guarantees that the intersection of the supports of $[s(U)]$ and $\mathrm{ch}(X, U)$ is compact. Therefore the intersection number $[s(U)] \cap \cap \mathrm{ch}(X, U)$ is well defined.

THEOREM. $IL(f, K) = [s(U)] \cap \mathrm{ch}(X, U)$.

Applying this theorem to the zero vector field, we get

COROLLARY [29]. *If $I\chi$ is the intersection homology Euler characteristic of X and $Z \subset T^*M$ is the zero section, $I\chi = [Z] \cap \mathrm{ch}(X)$.*

§ 8. Enter sheaf theory. The functor which assigns to each open set U in X the group $IC_i^{cl}(U)$ of intersection chains on U with closed support is a *sheaf.* This observation is key for axiomatic characterization of intersection homology as well as for most of its applications.

Because of the numbering conventions prevalent in sheaf theory, we define $\mathbf{IC}^i(U)$ to be $IC_{2n-i}^{cl}(U)$; if $U' \subset U$ then the map $\mathbf{IC}^i(U) \to \mathbf{IC}^i(U')$ is just restriction of geometric chains. Then $\mathbf{IC}^i$ is a sheaf because the allowability conditions of § 2 are local. The boundary map gives a map of sheaves $\delta: \mathbf{IC}^i \to \mathbf{IC}^{i+1}$ such that $\delta \circ \delta = 0$, so we have a complex of sheaves $\mathbf{IC}^\cdot$. As with any complex of sheaves, we can apply several cohomological functors. First is the cohomology sheaf functor $\mathbf{H}^i\mathbf{IC}^\cdot = \mathrm{Ker}\,\delta: \mathbf{IC}^i \to \mathbf{IC}^{i+1}/\mathrm{Im}\,\delta: \mathbf{IC}^{i-1} \to \mathbf{IC}$. This is a sheaf whose stalk $\mathbf{H}^i(\mathbf{IC}^\cdot)_p$ at p is $IH_{2n-i}(X, X-p)$. Second is the hypercohomology functor $\boldsymbol{H}^i(X, \mathbf{IC}^\cdot)$. Since $\mathbf{IC}^i$ is soft, for all i, the group $\boldsymbol{H}^i(X, \mathbf{IC}^\cdot)$ may be computed as the global section cohomology $\ker \delta\mathbf{IC}^i(X) \to \mathbf{IC}^{i+1}(X)/\mathrm{Im}\,\delta: \mathbf{IC}^{i-1}(X) \to \mathbf{IC}^i(X)$. Hence $\boldsymbol{H}^i(X, \mathbf{IC}^\cdot) = IH_{2n-i}^{cl}(X)$ which we also denote $IH^i(X)$. Third is hypercohomology with compact support $IH_c^i(X, \mathbf{IC}^\cdot)$ which for the same reason is $IH_{2n-i}(X)$.

If $\mathbf{P}^\cdot$ and $\mathbf{Q}^\cdot$ are two complexes of sheaves, a *quasi-isomorphism from $\mathbf{P}^\cdot$ to $\mathbf{Q}^\cdot$* is a diagram of complexes $\mathbf{P}^\cdot \xleftarrow{p} \mathbf{R}^\cdot \xrightarrow{q} \mathbf{Q}^\cdot$ so that p and q induce isomorphisms $\mathbf{H}^i\mathbf{P}^\cdot \xleftarrow{\cong} \mathbf{H}^i\mathbf{R}^\cdot \xrightarrow{\cong} \mathbf{H}^i\mathbf{Q}^\cdot$ for all i, or equivalently if for all open sets U, p and q induce isomorphisms $\boldsymbol{H}^i(U, \mathbf{P}^\cdot) \leftarrow \boldsymbol{H}^i(U, \mathbf{R}^\cdot) \to \boldsymbol{H}^i(U, \mathbf{Q}^\cdot)$.

If there is a quasi-isomorphism from $\mathbf{P}^{\cdot}$ to $\mathbf{Q}^{\cdot}$, then $\mathbf{P}^{\cdot}$ and $\mathbf{Q}^{\cdot}$ are called *quasi-isomorphic*. This is an equivalence relation. Quasi-isomorphic sheaves are interchangeable for all calculations with cohomological functors.

For any point p in X we choose a local analytic embedding of X near p in C^m and call B_p the intersection of a small open ball in C^m centered at p with X. It follows from stratification theory that the topology of B_p depends only on p. Straight forward geometric arguments show that the complex $\mathbf{IC}^{\cdot}$ satisfies the following four properties:

(0) *Boundedness and constructibility*: $\mathbf{IC}^i = 0$ if $i < 0$ or if i is large enough; and for some stratification $\{S_a\}$ of X, $\mathbf{H}^i\mathbf{IC}^{\cdot}\,|\,S_a$ is locally constant and finitely generated for all i for all a.

(1) *Support*: for all $i > 0$,

$$\dim_C\{x \in X \mid \ H^i(B_x; \mathbf{IC}^{\cdot}) \neq 0\} < n-i.$$

(2) *Cosupport*: For all $i > 0$,

$$\dim_C\{x \in X \mid \ H_c^{2n-i}(B_x; \mathbf{IC}^{\cdot}) \neq 0\} < n-i.$$

(3) *Normalization*: For some stratification $\{S_a\}$ of X, $\mathbf{H}^i\mathbf{IC}^{\cdot}/(X-\Sigma)$ is zero for $i \neq 0$ and is the constant sheaf for $i = 0$.

THEOREM [4]. *The sheaf* $\mathbf{IC}^{\cdot}$ *is uniquely characterized up to quasi-isomorphism by the above conditions* (0), (1), (2), *and* (3).

It is useful to consider also intersection homology with twisted coefficients. The coefficients will be a local system L (i.e. a locally constant sheaf) on $X-\Sigma$. The group $IH_i(X, L)$ may be defined as in § 2 or we may define the sheaf $\mathbf{IC}^i(L)$ directly as follows. The value of $\mathbf{IC}\ (L)$ on an open set $U \subset X$ is those $2n-i$ chains c with closed supports on $U \cap (X-\Sigma)$ with coefficients in L satisfying the allowability condition: the closure in U of the support of c intersects each singular stratum $S_a \subset \Sigma$ in a set of real dimension less than $(2n-i)-c_a$ and the closure in U of the support of the boundary ∂c intersects each singular stratum $S_a \subset \Sigma$ in a set of real dimension less than $(2n-i-1)-c_a$. The sheaf $\mathbf{IC}^{\cdot}(L^{\cdot})$ may be characterized as in the above theorem replacing the normalization condition (3) by

(3′) *Normalization*: $\mathbf{H}^i\mathbf{IC}^{\cdot}(L)|X-\Sigma$ is zero for $i \neq 0$ and is L for $i = 0$.

Conjecturally, if L is a polarized variation of Hodge structure then $H^i(\mathbf{IC}^{\cdot}(L)) = IH^i(X, L)$ has a pure Hodge structure. This is verified for curves [82].

Sheaf theory enables one to give local expressions for some basic properties of intersection homology. For example Poincaré duality of § 2 becomes the statement that $\mathbf{IC}^\cdot(Q)$ is quasi-isomorphic to its dual (see [16] or [75]).

§ 9. Perverse sheaves. Intersection homology sheaves are objects in a beautiful Abelian category called the category $P(X)$ of perverse sheaves [12].

The category C of complexes of sheaves on X is deficient from the point of view of homological functors because quasi-isomorphisms $\mathbf{P}^\cdot \leftarrow \mathbf{R}^\cdot \rightarrow \mathbf{Q}^\cdot$ may not be morphisms in C, and even if they are (when $\mathbf{P}^\cdot \overset{=}{\leftarrow} \mathbf{P}^\cdot \rightarrow \mathbf{Q}^\cdot$) they may not be invertible in C. This situation may be remedied by introducing formal inverses of morphisms in C which are quasi-isomorphisms [46], [34]. The resulting category $D(X)$ is called the *derived category* of the category of sheaves on X. The isomorphisms in $D(X)$ are exactly the quasi-isomorphisms. But $D(X)$ is not Abelian: instead it has the structure of a "triangulated category" which is quite complicated [77].

The category $P(X)$ of perverse sheaves on X is the full subcategory of $D(X)$ whose objects satisfy the following slight weakening of the conditions characterizing the intersection homology sheaf $\mathbf{IC}^\cdot$:

DEFINITION. A complex of sheaves $\mathbf{K}^\cdot$ on X is called a *perverse sheaf* if it satisfies the following three properties:

(0) *Boundedness and constructibility*: $\mathbf{K}^\cdot = 0$ if $i < 0$ or if i is large enough; and for some stratification $\{S_a\}$ of X, $\mathbf{H}^i\mathbf{K}^\cdot|S_a$ is locally constant and finitely generated for all i and for all a.

(1) *Support*: For all i,

$$\dim_C\{x \in X \mid \boldsymbol{H}^i(B_x; \mathbf{IC}^\cdot) \neq 0\} \leqslant n-i.$$

(2) *Cosupport*: For all i,

$$\dim_C\{x \in X \mid \boldsymbol{H}_c^{2n-i}(B_x; \mathbf{IC}^\cdot) \neq 0\} \leqslant n-i.$$

We now digress to show how intersection homology sheaves provide a rich supply of perverse sheaves.

DEFINITION. An *enriched subvariety* of X is a pair $(V, \boldsymbol{L})$ where V is locally closed, non-singular, equidimensional subvariety and $\boldsymbol{L}$ is a local system of coefficients on V. Two enriched subvarieties $(V, \boldsymbol{L})$ and $(V', \boldsymbol{L}')$ are considered equal if $V \cap V'$ is dense in V and in V', and $\boldsymbol{L}|(V \cap V') = \boldsymbol{L}'|(V \cap V')$. An *irreducible* enriched subvariety $(V, \boldsymbol{L})$ is one where V is an irreducible variety and $\boldsymbol{L}$ is an irreducible local system on V. An enriched subvariety $(V, \boldsymbol{L})$ gives rise to a complex $\mathbf{IC}^{\cdot}(\overline{V}, \boldsymbol{L})$ on X called the intersection homology sheaf of $(V, \boldsymbol{L})$ by extending the complex $\mathbf{IC}^{\cdot}(\boldsymbol{L})$ on $\overline{V}$ by zero in X.

If $\mathbf{A}^{\cdot}$ is a complex of sheaves and c is an integer, we define $\mathbf{A}^{\cdot}[c]$ to be the same complex with the numbering shifted by c. In other words, the ith sheaf in $\mathbf{A}^{\cdot}[c]$ is $\mathbf{A}^{i+c}$.

Comparison of the definition of a perverse sheaf with the theorem of § 8 shows that if c is the codimension of V in X, $\mathbf{IC}^{\cdot}(\overline{V}, \boldsymbol{L})[-c]$ is a perverse sheaf on X. For example, if $x \in V$, we have $\boldsymbol{H}^{c_a}(B_x; \mathbf{IC}^{\cdot}(V, \boldsymbol{L})[-c]) = \boldsymbol{L}_x$ and $\boldsymbol{H}_c^{2n-c_a}(B_x; \mathbf{IC}^{\cdot}(V, \boldsymbol{L})[-c]) = \boldsymbol{L}_x$.

THEOREM. *The category $P(X)$ of perverse sheaves on X is an Artinian Abelian category whose simple objects are the complexes* $\mathbf{IC}^{\cdot}(\overline{V}, \boldsymbol{L})[-c]$ *for irreducible enriched subvarieties $(V, \boldsymbol{L})$ of X.*

The category $P(X)$ is important because of its applications (§ 11, §13 and § 16). It would be interesting to understand its structure more directly (see [60] and [35]).

Applications to analysis

§ 10. L^2 cohomology. The L^2 cohomology of the nonsingular part $X-\Sigma$ of any compact analytic variety X, provided with an appropriate polyhedral metric, was found to be finite-dimensional and to satisfy Poincaré duality [24], [25]. This led to the question, resolved affirmatively, of whether this L^2 cohomology was in fact intersection homology with real coeficients. In this section we address the same question for metrics more naturally associated with the analytic structure of X.

We define $\Omega^i_{(2)}(X-\Sigma)$ to be the space of smooth i forms ω on $X-\Sigma$ such that

$$\int_{X-\Sigma} \omega \wedge *\omega < \infty, \qquad \int_{X-\Sigma} d\omega \wedge *d\omega < \infty.$$

We define the L^2 cohomology of $X-\Sigma$, $H^i_{(2)}(X-\Sigma)$, to be the ith cohomology of the complex $\Omega^{\cdot}_{(2)}(X-\Sigma)$. Since the operator $*$ depends on the Riemannian metric chosen on $X-\Sigma$, the group $H^i_{(2)}(X-\Sigma)$ also depend on this metric. We are interested in two contexts which differ by the choice of this metric:

Context I. The variety X is embedded analytically in a Kähler manifold M; the metric chosen on $X-\Sigma$ is induced via this embedding from the Kähler metric on M.

Context II. The manifold $X-\Sigma$ with its metric has finite volume and is obtained as the quotient space of a discrete group acting properly on a Hermitian symmetric domain; X is the Baily–Borel compactification of $X-\Sigma$.

The following conjectures may be summarized by the philosophy that analysis on $\Omega_{(2)}(X-\Sigma)$ should be related to analysis on $\tilde{X}$ where $\tilde{X}\to X$ is a small resolution (see § 5).

CONJECTURE. $H^i_{(2)}(X-\Sigma)\cong IH_{2n-i}(X;\boldsymbol{R})$ (see [26] for context I and [83] for context II).

CONJECTURE. Every class in $H^i_{(2)}(X-\Sigma)$ contains a unique harmonic (closed and coclosed) representative. (This is well-known in context II because the metric is complete.) A Hodge theory can be obtained by splitting the L^2 harmonic forms with values in $\boldsymbol{C}$ into their (p,q) pieces and it satisfies the properties of the conjecture of § 4.

CONJECTURE. The index of the $L^2\bar{\partial}$ complex is the arithmetic genus of $\tilde{X}$ where $\pi\colon \tilde{X}\to X$ is any resolution of singularities of X. More generally, the $K_*(X)$ element obtained from this complex by the analytic procedure of [8] coincides with the π pushforward of the K orientation of $\tilde{X}$ [10].

Work on these conjectures has been difficult for lack of adequate analytic methods to study $\Omega^{\cdot}_{(2)}(X-\Sigma)$ near a singularity of X and for lack of adequate information on the metric structure of X near a singularity. Nevertheless, partial results are known. The first holds in context I for metrically conical singularities [25] and for some surfaces [49] and in context II for the case where Σ is a manifold [15] and for some rank 2 cases [84], [23]. The third holds in context I for curves [47].

§ 11. Relation to algebraic analysis. Holonomic $\mathfrak{D}$-modules are an important and beautiful subject in their own right, but here we treat them only in relation to our topological constructions. Suppose M is a non-singular analytic variety. Let $\mathfrak{D}$ be the sheaf of linear differential operators with analytic coefficients. Then $\mathfrak{D}$, as a sheaf of noncommutative rings on M, is filtered by the order. The associated graded $\mathrm{Gr}\,\mathfrak{D}$ is commutative; sections of it over $U\subset M$ may be interpreted as functions on T^*U. Suppose $\mathfrak{M}$ is a coherent sheaf of $\mathfrak{D}$-modules. $\mathfrak{M}$ is called *holonomic r.s.* (for regular singularities) if it has a filtration as a module over filtered $\mathfrak{D}$ so that $\mathrm{Gr}\,\mathfrak{M}$ has support in T^*M which is reduced and of pure dimension m. (This definition provides no intuitive feeling for holonomic r.s. $\mathfrak{D}$-modules; for this we refer the reader to [66], [13].) We denote by $\mathrm{ch}(M)$ (for characteristic variety) the algebraic cycle in $T^*(M)$ determined by $Gr\,\mathfrak{M}$. Any $\mathfrak{D}$-module $\mathfrak{M}$ determines a DeRham complex $\mathrm{DR}\,\mathfrak{M} = \mathfrak{M}\otimes\Omega^0 \xrightarrow{d} \mathfrak{M}\otimes\Omega^1 \xrightarrow{d}$ in the derived category $D(M)$ (see § 8).

THEOREM ([63], [50], [18]). *The category of holonomic r.s. $\mathfrak{D}$-modules on M is equivalent to the category $P(M)$ of perverse sheaves on M. The equivalence is given by the DeRham functor* DR.

Therefore the irreducible holonomic r.s. $\mathfrak{D}$-modules correspond to intersection homology sheaves of subvarieties of M. If $\mathrm{DR}\,\mathfrak{L} = \mathbf{IC}^\cdot(X)$ where $X\subset M$, then $\mathrm{ch}\,\mathfrak{L} = \mathrm{ch}(X)$ as defined in § 6.

There is a filtration of the $\mathfrak{D}$-module $\mathfrak{L}$ such that $\mathrm{DR}\,\mathfrak{L} = \mathbf{IC}^\cdot(X)$ which conjecturally gives rise to the Hodge filtration suggested in § 4 on $IH_i(X)$ [18]. The Fourier transformation, taking the category of radially homogeneous holonomic r.s. $\mathfrak{D}$-modules on $\boldsymbol{C}^n$ to itself, leads to a similar operation on $P(\boldsymbol{C}^n)$ with interesting applications [22], [19].

Applications to algebraic geometry

§ 12. The decomposition theorem. The decomposition theorem says that the pushforward of an intersection homology sheaf by a proper algebraic map is a direct sum of intersection homology sheaves. It contains as special cases the deepest homological properties of algebraic maps that we know. It was conjectured in [36] and proved in [12].

THEOREM. (1) *Let $f\colon X\to Y$ be a proper projective map of complex algebraic varieties. Then there exist a unique set of irreducible enriched subvarieties $\{(V_a, \boldsymbol{L}_a)\}$ in Y (V_a smooth in Y; $\boldsymbol{L}_a$ a local system of $\boldsymbol{Q}$ vector spaces on V_a) and polynomials $\{\varphi^a = \varphi_0^a+\varphi_1^a t+\dots\}$ such that there is a quasi-isomorphism*

$$f_*\mathbf{IC}^\cdot(X) \approx \bigoplus_{a,i} \mathbf{IC}^\cdot(\overline{V}_a, \boldsymbol{L}_a)[-i]\otimes \boldsymbol{Q}^{\varphi_i^a}. \tag{**}$$

(2) *The coefficients of φ^a are palindromic around $k_a = \dim X-\dim V_a$ (i.e. $t^{k_a}\varphi^a(t^{-1}) = \varphi^a(t)$) and the odd and even degree terms are separately unimodal (i.e. if $i\leqslant k_a$ then $\varphi_{i-2}^a\leqslant\varphi_i^a$).*

Applying hypercohomology to the above quasi-isomorphism, we get

$$P^X(t) = \sum_a \varphi^a(t)P^a(t),$$

where P^X is the Poincaré polynomial for $IH^*(X)$ and P^a is the Poincaré polynomial for $IH^*(\overline{V}_a, L_a)$. If X is compact, the $\overline{V}_a$ will also be compact so P^X and P^a have the character of the Poincaré polynomial of a smooth projective variety. Part (2) of the theorem asserts that the φ^a also have this character (palindromicity $\approx$ Poincaré duality and unimodality $\approx$ hard Lefschetz). So the theorem asserts that $IH^*(X)$ is a sum of terms, each of which is like the intersection homology of the product of a fictitious fiber variety (with Poincaré polynomial φ^a), and an enriched subvariety $(V_a, \boldsymbol{L}_a)$ of Y. (Conjecturally, $\boldsymbol{L}_a$ should be a polarized variation of Hodge structure.)

The quasi-isomorphism in the decomposition formula (**) can be made canonical by the following procedure [28] given a factorization of f as $X \hookrightarrow Y \times \boldsymbol{P}^m \to Y$. The hyperplane class $[H]$ in $H^2(\boldsymbol{P}^m)$ induces a map η: $f_*\mathbf{IC}^\cdot(X) \to f_*\mathbf{IC}^\cdot(X)$ [2] essentially by transversally intersecting cycles with $[H]$. If we denote by $\mathbf{A}^\cdot(i)$ the sum of the right-hand side of (**) over a so (**) reads $f_*\mathbf{IC}^\cdot(X) \approx \bigoplus_i \mathbf{A}^\cdot(i)$, then η decomposes into pieces η_{ij}: $\mathbf{A}^\cdot(i) \to A^\cdot(i-j)$ [2]. Set $\eta_j = \sum_i \eta_{ij}$. Then there is a unique quasi-isomorphism in (**) satisfying $(\mathrm{ad}\,\eta_0)^{i-1}\eta_i = 0$ for all i (where $\mathrm{ad}\,\eta_0\xi$ is $\eta_0\xi - \xi\eta_0$).

If $X \to Y$ is a resolution of singularities of Y, then one of the enriched subvarieties in the decomposition will be $(Y, \boldsymbol{Q})$ and the corresponding φ^a will be 1. Thus $IH^*(Y)$ sits canonically in the cohomology of X given a factorization of the resolution as in the last paragraph. Conjecturally, this embedding determines the Hodge structure on $IH^*(Y)$ from that on X.

The decomposition theorem contains, for example, the invariant cycle theorem and the degeneration of the spectral sequence for f in case f is a topological fibration. See [40] for a discussion of some of its consequences. It is also one of the most powerful techniques for calculating intersection homology (see [5]). At the moment, the only proof of it goes through characteristic p algebraic geometry. (As a consequence, it is unproved for proper analytic maps.) It would be very interesting to find an analytic proof, either using $\mathfrak{D}$-modules or using L^2 techniques.

§ 13. Specialization. Suppose $\pi: X \to C$ is a map of an algebraic variety to a smooth algebraic curve, o is a point in C, and $c \in C$ is a nearby general point (i.e. π is a topological fibration near c). There is a continuous map called ψ from X_c, the fiber over c, to X_o the fiber over o, which roughly collapses points near a given stratum S in X_o to S [41]. There is also a map μ: $X_c \to X_c$ with $\psi \circ \mu = \psi$ called the *monodromy* which represents the result of tracing paths over c as it moves once around o. Therefore there is a complex of sheaves $\psi_* \mathbf{IC}^{\cdot}(X_c)$ on X_0 called the *nearby cycles* of $\mathbf{IC}^{\cdot}$ with an action of μ_* on it. Although there is some choice in the specification of ψ and μ, the complex $\psi_* \mathbf{IC}^{\cdot}(X_c)$ and the action of μ_* are independent up to quasi-isomorphism of the choice. (Moreover, they can be defined purely algebraically [69].)

PROPOSITION ([41], [12]). *The complex* $\psi_* \mathbf{IC}^{\cdot}$ *is a perverse sheaf.*

Therefore we dispose of the Abelian category structure of $P(X_0)$ to analyze the monodromy μ_*. There is a factorization $\mu_* = F \cdot (1+N)$, where F has finite order and N is nilpotent. Then there is a unique increasing filtration W^i on $\psi_* \mathbf{IC}^{\cdot}$ such that N sends W^i to W^{i-2} and N^i takes $\mathrm{Gr}^i \mathbf{IC}^{\cdot}$ isomorphically to $\mathrm{Gr}^{-i} \mathbf{IC}^{\cdot}$, where Gr^i is the associated graded to the filtration W.

THEOREM ([33]). *The graded pieces* $\mathrm{Gr}^i \psi_* \mathbf{IC}^{\cdot}(X_c)$ *are semi-simple in* $P(X_0)$. *In other words, they are direct sums of intersection homology sheaves of irreducible enriched subvarieties of* X_0.

The study of $\psi_* \mathbf{IC}^{\cdot}$ can be generalized to the case where X_0 is replaced by an arbitrary hypersurface [78].

Applications to group theory

§ 14. Weyl group representations. Here and in the succeeding sections, we will treat only the case of $\mathrm{GL}(k, \boldsymbol{C})$. Lie theorists can imagine the generalization to an arbitrary reductive complex algebraic group with the aid of occasional parenthetical remarks.

Consider the variety $\mathfrak{N}$ of all $k \times k$ complex matrices all of whose eigenvalues are zero. It is singular, and it has a stratification $\{S_a\}$ indexed by partitions a of the number k. The stratum S_a consists of matrices whose sizes of Jordan blocks are given by a. Let c_a be the codimension of S_a in $\mathfrak{N}$. The variety $\mathfrak{N}$ also has a resolution π: $\tilde{\mathfrak{N}} \to \mathfrak{N}$ constructed as follows: Let $\mathfrak{B}$ denote the manifold of complete flags $0 \subset F_1 \subset F_2 \subset \ldots \subset F_{k-1} \subset \boldsymbol{C}^k$ in $\boldsymbol{C}^k$. Then $\tilde{\mathfrak{N}}$ is the subvariety of $\mathfrak{N} \times \mathfrak{B}$ consisting of pairs $(x, \{F_j\})$ so that $xF_j \subset F_{j-1}$. The map π is the projection to the first factor. The decomposition formula (§ 12) for $\mathbf{A}^{\cdot} = \pi_* \mathbf{IC}^{\cdot}(\tilde{\mathfrak{N}})$ may be computed as follows:

PROPOSITION ([17]).

$$\mathbf{A}^{\cdot} = \bigoplus_a \mathbf{IC}^{\cdot}(\bar{S}_a)[-c_a] \otimes \boldsymbol{Q}^{\varphi^a}.$$

So $\mathbf{A}^{\bullet}$ is a semisimple perverse sheaf on $\mathfrak{N}$. (For other reductive groups, enriched strata will be necessary.)

The symmetric group on k letters, denoted W (the Weyl group of $\mathrm{GL}(k)$), acts on $\mathbf{A}^{\bullet}$ by automorphisms in $P(\mathfrak{N})$, the category of perverse objects on $\mathfrak{N}$. This action, constructed first in [72], has several descriptions; we follow [71]. We map $\mathfrak{G}$, the space of all $k\times k$ complex matrices, to C^k by the coefficients of the characteristic polynomial. Then $\mathfrak{N}$ is the fiber over 0, and if $c\in C^k$ is a nearby non-singular point, it turns out that $\mathbf{A}^{\bullet}$ is $\psi_*\mathbf{IC}^{\bullet}(\mathfrak{G}_c)$. Just as in § 13, the fundamental group of the complement of the discriminant in C^k acts on $\mathbf{A}^{\bullet}$ by monodromy. This fundamental group is the braid group on k strands; its action factors through the quotient map to W.

PROPOSITION ([17]). *The action of W on $\mathbf{A}^{\bullet}$ induces an isomorphism from the group ring of W to the endomorphism ring of $\mathbf{A}^{\bullet}$ in $P(\mathfrak{N})$.*

It follows that isotypical components $\mathbf{IC}^{\bullet}(\bar{S}_a)[-c_a)\otimes Q^{v^a}$ of $\mathbf{A}^{\bullet}$ correspond bijectively to irreducible representations of W. This correspondence between partitions a and irreducible W representations P_a is that of Young. Applying hypercohomology, we obtain as another corollary the formula that the multiplicity of p_a in the standard action of W on $H^i(\mathfrak{B})$ is $\dim IH^{i-c_a}(\bar{S}_a)$ ([17], [54]).

§ 15. Representations of Hecke algebras. We sketch the contents of the papers [51], [52], whose historical importance in stimulating the recent development of the material of this report cannot be overemphasized. The form of our presentation relies on these later developments.

The variety $\mathfrak{B}\times\mathfrak{B}$, where $\mathfrak{B}$ is the flag manifold as in § 14, is stratified by orbits S of the diagonal action of $\mathrm{GL}(k, C)$. The orbits are parametrized by elements a of the symmetric group W. A pair of flags $\{F_j\}$, $\{F'_j\}\in\mathfrak{B}\times\mathfrak{B}$ corresponds to $a\in W$ if there is a basis $e_1, e_2, \ldots, e_k$ of C^k so that F_j is the span of $e_1, e_2, \ldots, e_j$ and F'_j is the span of $e_{a(1)}, e_{a(2)}, \ldots, e_{a(j)}$.

The Hecke algebra $\mathfrak{H}$ is the set of formal linear combinations of intersection homology sheaves $\sum\limits_{a,i} m_{a,i}\mathbf{IC}^{\bullet}(\bar{S}_a)[i]$ with integral multiplicities $m_{a,i}$. Addition is addition of multiplicities; multiplication is obtained by regarding the intersection homology sheaves as "sheaf valued correspondences" on $\mathfrak{B}\times\mathfrak{B}$. That is $\mathbf{A}^{\bullet}\cdot\mathbf{B}^{\bullet} = p_{13*}(p_{12}^*\mathbf{A}^{\bullet}\otimes p_{23}^*\mathbf{B}^{\bullet})$ where p_{13}, p_{12}, and p_{23} map $\mathfrak{B}\times\mathfrak{B}\times\mathfrak{B}$ to $\mathfrak{B}\times\mathfrak{B}$ by forgeting the factor not named. The Hecke algebra is an algebra over $\boldsymbol{Z}[t, t^{-1}]$ where t sends $\mathbf{IC}^{\bullet}(\bar{S}_a)[i]$ to $\mathbf{IC}^{\bullet}(\bar{S}_a)[i-1]$. (For general reductive Lie groups, $\mathfrak{H}$ depends only on the Weyl group W, even though $\mathfrak{B} \times \mathfrak{B}$ and the singularities of the strata are not determined by W alone.)

There is a positive cone $\mathfrak{K} \subset \mathfrak{H}$ consisting of actual complexes of sheaves on $\mathfrak{B} \times \mathfrak{B}$, i.e. $\mathfrak{K}$ consists of formal linear combinations where all multiplicities $m_{a,i}$ are non-negative. Clearly $\mathfrak{K}$ is closed under multiplication in $\mathfrak{H}$. This defines a partial ordering $\leqslant_L$ of the elements of W by $a \leqslant_L a'$ if $\mathfrak{K} \cdot \mathbf{IC}^\cdot(\bar{S}_a)$ contains an element where $\mathbf{IC}^\cdot(\bar{S}_{a'})[i]$ occurs with positive multiplicity. A *left cell* in W is a $\leqslant_L$ equivalence class of elements of W. Any left cell $C \subset W$ gives rise to a representation $R(L)$ of the Hecke algebra $\mathfrak{H}$ with C as a $\mathbf{Z}[t, t^{-1}]$ basis as follows $R(C)$ is a subquotient of the regular representation of $\mathfrak{H}$; say $R(C) = \mathfrak{S}/\mathfrak{T}$. Here $\mathfrak{S} \subset \mathfrak{H}$ is generated over $\mathbf{Z}[t, t^{-1}]$ by $\mathbf{IC}^\cdot(\bar{S}_a)$ where $\{C\} \leqslant_L a$ and $\mathfrak{T}$ is generated by $\mathbf{IC}^\cdot(\bar{S}_a)$, where $\{C\} <_L a$.

THEOREM. *The representations $R(C)$ are irreducible. All irreducible representations of $\mathfrak{H}$ occur as $R(C)$ for some left cell C.*

(This theorem does not generalize easily to other reductive groups.) The main interest in this theorem is that the combinatorics of it can be explicitly spelled out. The left cells C arise from partitions of k by the Robinson–Schensted algorithm. The representations $R(C)$ may be written with respect to the basis $\{\mathbf{IC}^\cdot(S_a), a \in L\}$ using inductively computable combinatorial objects called W-graphs. If t is specialized to 1, then $\mathfrak{H}$ becomes the group ring $\mathbf{Z}[W]$ of W and the theorem gives the irreducible representations of the symmetric group W with a basis with particularly agreeable properties: for example, all elements are represented by matrices with integral entries.

The methods of this section can be used to compute both the local and global intersection homology groups of the Schubert varieties $\bar{S}_a$. A fascinating phenomenon is that these groups are all zero in odd degrees. The same holds for the computations for toric varieties [5], nilpotent varieties [17], and K_C orbits in $\mathfrak{B}$ [57]. One would like a general explanation for this phenomenon.

§ 16. Lie algebra representations. In this section, we discuss the serendipitous entry of intersection homology into the study of infinite-dimensional representations of Lie algebras.

Recall that the $k \times k$ complex matrices under the bracket operation, denoted $\mathfrak{G}$, is the Lie algebra of the group of invertible $k \times k$ matrices under composition, denoted G. The sub-Lie algebra $\mathfrak{n}^+$ of upper triangular matrices that are zero on the diagonal is the Lie algebra of the subgroup

N^+ of upper triangular matrices that are one on the diagonal. A $(\mathfrak{G}, N^+)$ representation V is a possibly infinite-dimensional representation of $\mathfrak{G}$ such that the action of $\mathfrak{n}^+$ exponentiates to representation of N^+ and, for all vectors $v \in V$, N^+v is contained in a finite-dimensional subspace on which N acts algebraically. In order to focus on the part of the theory where the transition from algebra to topology is fully understood, we will consider only $\bar{\mathcal{O}}_1$ the category of $(\mathfrak{G}, N^+)$ representations V satisfying the following condition:

The space V has a finite decomposition into summands V_i on each of which the center Z of $U(\mathfrak{G})$, acts as it does on some irreducible finite-dimensional representation ϱ_i of $\mathfrak{G}$.

Recall that $U(\mathfrak{G})$ is the universal enveloping algebra of $\mathfrak{G}$. It is the associative algebra containing $\mathfrak{G}$ such that the embedding induces an equivalence between the representation theory of $\mathfrak{G}$ and that of $U(\mathfrak{G})$. Every element of its center acts on a finite-dimensional irreducible representation ϱ by a scalar.

The passage to topology proceeds in four steps:

Step I. The category $\bar{\mathcal{O}}$ is the direct sum of the categories $\bar{\mathcal{O}}_\varrho$ for ϱ a finite-dimensional irreducible representation of $\mathfrak{G}$, defined by replacing $*$ by the condition that Z acts on all of V as it does on ϱ. So it suffices to understand $\bar{\mathcal{O}}_\varrho$.

Step II. The categories $\bar{\mathcal{O}}_\varrho$ are all equivalent to each other by an algebraic process called coherent continuation [14]. So it suffices to understand $\bar{\mathcal{O}}_1$ where 1 is the trivial one-dimensional representation.

Step III [21], [11]. The category $\bar{\mathcal{O}}_1$ is equivalent to the category of holonomic r.s. $\mathfrak{D}$-modules on the flag manifold $\mathfrak{B}$ constructible with respect to the stratification $\{S_a\}$ of $\mathfrak{B}$ by N^+ orbits. (Saying that a holonomic r.s. $\mathfrak{D}$-module $\mathfrak{M}$ is constructible with respect to $\{S_a\}$ means that its characteristic variety $\mathrm{ch}(\mathfrak{M})$ is contained in the union of the conormal bundles $C(S_a)$. The strata S_a are Schubert cells: they are restrictions to a slice point $\times\mathfrak{B}$ of the strata of $\mathfrak{B}\times\mathfrak{B}$ of § 15. So the S_a are again naturally indexed by the symmetric group W.) The equivalence is given by associating to a $\mathfrak{D}$-module $\mathfrak{M}$ the vector space of its global sections. This is a representation of $\mathfrak{G}$ because elements of $\mathfrak{G}$ give vector fields on $\mathfrak{B}$ which are global sections of $\mathfrak{D}$.

Step IV. As in § 9 the category of holonomic r.s. $\mathfrak{D}$-modules on $\mathfrak{B}$, constructible with respect to the strata S_a is equivalent to the category of perverse sheaves on $\mathfrak{B}$ constructible with respect to S_a. (A perverse sheaf $\mathbf{A}^\bullet \in P(\mathfrak{B})$ is constructible with respect to $\{S_a\}$ if all of its homology sheaves $\mathbf{H}^i(\mathbf{A}^\bullet)$ are locally constant on the S_a.)

With the transition completed, we see that irreducible representations in $\bar{\mathcal{O}}$ correspond to intersection homology sheaves $\mathbf{IC}^{\bullet}(\bar{S}_\alpha)[-c_\alpha]$. Any purely category-theoretic question about $\bar{\mathcal{O}}$ can be answered topologically using $P(\mathfrak{B})$. As an example (the historically motivating one), the Kazhdan–Lusztig conjectures [51], were proved this way [11], [21]. A Verma module in $\bar{\mathcal{O}}$ is a representation that is $U(\mathfrak{n}^-)$ free where $\mathfrak{n}^-$ is the lower triangular matrices zero on the diagonal. It corresponds in $P(\mathfrak{B})$ to a sheaf whose stalk Euler characteristic is 1 on one S_β and 0 on all the other strata. If L is the irreducible module in $\bar{\mathcal{O}}_p$ corresponding to $\mathbf{IC}^{\bullet}(\bar{S}_\alpha)[-c_\alpha]$, it has a resolution $\ldots V_2 \to V_1 \to V_0 \to L$ where the V_i are direct sums of Verma modules. The Kazhdan–Lusztig conjectures state that $\sum(-1)^i m_i$ where m_i is the multiplicity of the Verma module in $\bar{\mathcal{O}}_p$ corresponding to S, is the Euler characteristic of the stalk homology of $\mathbf{IC}^{\bullet}(\bar{S}_\alpha)[-c_\alpha]$ at a point in S_β. Given all of the above facts, the proof is an exercise.

There is an extension of the above theory to $(\mathfrak{G}, K)$ modules whenever K has only finitely many orbits on $\mathfrak{B}$ [11], [79]. This applies to Harish-Chandra modules. There are also two other similar but unproved conjectured relations between algebra and topology: one for representations of p-adic groups [81] and one for modular representations [55].

§ 17. Other subjects. Only some subjects relating both to topology and to complex algebraic geometry have been treated. This leaves out much interesting work on intersection homology in the two fields separately.

In topology, there is an L class in $H_*(X)$ which relates to the intersection homology signature [3], [24]. There is a singular cobordism theory for a class of spaces called rational Witt spaces which are the most general on which rational Poincaré duality holds. The cobordism groups are the Witt ring of Q in dimension $4n$ and 0 otherwise [70]. There is an integral version of this [65] and an application of these ideas to the proof of the *Hauptvermutung* [64]. There is a theory of intersection homology operations [38], and a theory of obstructions to immersion [44].

In algebraic geometry, there is a construction of $\mathbf{IC}^{\bullet}(X)$ in the l-adic topology for varieties in characteristic p. If X is complete, Frobenius F acts on $IH^i(X)$ with eigenvalues of absolute $p^{i/2}$, so statements about the number of points defined over $F(p^k)$ analogous to those of the Weil conjectures are obtained [27], [12]. For an algebraic group G defined over $F(p^k)$, there is a collection of intersection homology sheaves on G such that interesting characters of G are evaluated at $p \in G$ by taking the alternating trace of a particular automorphism of the sheaves in the stalk over p [56].

We close with a general remark on the enterprise of studying the global topology of singular spaces. Usually a given concept in the topology of manifolds such as homology or a characteristic class has several plausible extensions to singular spaces. The art is to find the most useful one. When this is done, it often happens that one finds new results about the non-singular case. For example, results like those of [12] and [13] would not have been anticipated five years ago, even when all spaces involved are smooth.

Bibliography

General references on intersection homology

[1] *Bern Seminar on Intersection Homology* (A. Borel, ed.), to appear.
[2] Brylinski J.-L., (Co)-homologie d'intersection et faisceaux pervers, Sem. Bourbaki No. 585, *Astérisque* **92–93** (1982), pp. 129–158.
[3] Goresky M. and MacPherson R., Intersection Homology Theory, *Topology* **19** (1980), pp. 135–162.
[4] Goresky M. and MacPherson R., Intersection Homology II, *Inventiones Math.* **72** (1) (1983), pp. 77–130.
[5] MacPherson R., *Intersection Homology,* Hermann Weyl lectures, to appear.
[6] Proceedings of the conference *Analyse et topologie sur les espaces singuliers,* juillet 1981, to appear in *Astérisque.*
[7] Springer T., Quelques applications de la cohomologie d'intersection, Sem. Bourbaki No. 589, *Astérisque* **92–93** (1982), pp. 249–274.

Other references

[8] Baum P. and Douglas R., K Homology and Index Theory, *Proc Symp. in Pure Math* **38**, Part 1 (1982), pp. 117–173.
[9] Baum P., Fulton W., and MacPherson R., Riemann–Roch for Singular Varieties, *Publ. Math. I.H.E.S.* **45** (1976).
[10] Baum P., Fulton W., and MacPherson R., Riemann–Roch and Topological K-Theory for Singular Varieties, *Acta Math.* **143** (1979), pp. 155–192.
[11] Beilinson A. and Bernstein J., Localisation des $\mathfrak{g}$-modules, *C.R.A.S.* **292** (1981), pp. 15–18.
[12] Beilinson A., Bernstein J., and Deligne P., to appear in [6], above.
[13] Bernstein J., Algebraic Theory of $\mathfrak{D}$-Modules, Proc. of *$\mathfrak{D}$-modules et singularités,* to appear in *Astérisque.*
[14] Bernstein J. and Gelfand S., Tensor Products of Finite and Infinite-Dimensional Representations of Semisimple Lie Algebras, *Compositio Math.* **41** (1980), pp. 245–285.
[15] Borel A., L^2 Cohomology and Intersection Cohomology of Certain Arithmetic Varieties, *Proc. of E. Noether Symposium at Bryn Mawr,* Springer-Verlag, to appear.
[16] Borel A. and Moore J. C., Homology Theory for Locally Compact Spaces, *Michigan Math. J.* **7** (1960), pp. 137–159.

[17] Borho W. and MacPherson R., Représentations des groupes de Weyl et homologie d'intersection pour les variétés de nilpotents, *C. R. Acad, Sc. Paris* **292** (1981), pp. 707–710.

[18] Brylinski J.-L., Modules Holonomes à singularités régulières et filtration de Hodge I, *Algebraic Geometry (Proceedings La Rábida)*, Lect. Notes in Math. **961**, Springer-Verlag (1982), pp. 1–21.

[19] Brylinski J.-L., Transformations canoniques, dualité projective, transformation de Fourier, et sommes trigonometriques, to appear in *Astérisque*.

[20] Brylinski J.-L., Dubson A., and Kashiwara M., Formule de l'indice pour les modules holonomes et obstruction d'Euler locale, *C.R.A.S.* (1981).

[21] Brylinski J.-L. and Kashiwara M., Kazhdan–Lusztig Conjecture and Holonomic Systems, *Invent. Math.* **64** (1981), pp. 387–410.

[22] Brylinski J.-L., Malgrange B., and Verdier J.-L., Transformations de Fourier Topologique, to appear in *Comptes Rendus A. S. Paris*.

[23] Casselman W., *A Hodge Theory for Local L^2 Cohomology of Baily–Borel Compactifications, of rank two*, to appear.

[24] Cheeger J., On the Spectral Geometry of Spaces with Cone-Like Singularities, *P.N.A.S.* **76** (5) (1979), pp. 2103–2106.

[25] Cheeger J., On the Hodge Theory of Riemannian Pseudo-Manifolds, *Proc. Symposia Pure Math.* **36**, Providence, AMS (1980), pp. 91–146.

[26] Cheeger J., Goresky M., and MacPherson R., L^2-Cohomology and Intersection Homology for Singular Algebraic Varieties. In: S. Yau (ed.), *Proceedings of Year in Differential Geometry*, I.A.S. (1981), Annals of Math. Studies, Princeton.

[27] Deligne P., *Pureté de la cohomologie de MacPherson–Goresky*, d'après un exposé de O. Gabber, rédigé par P. Deligne, IHES preprint, Fév. 1981.

[28] Deligne P., Applications de la Pureté, Proc. of *$\mathfrak{D}$-modules et singularités*, to appear in *Astérisque*.

[29] Dubson A., to appear.

[30] Séminaire Ecole Normale Supérieure 1974–75, *Astérisque* **36–37** (1976).

[31] Séminaire Ecole Normale Supérieure (1978–79), *Astérisque* **82–83** (1981).

[32] Fulton W. and MacPherson R., Categorical Framework for the Study of Singular Spaces, *Memoirs of the Amer. Math. Soc.* **243**, A.M.S., Providence, RI (1981).

[33] Gabber O., to appear.

[34] Gabriel P. and Zismann M., *Calculus of Fractions and Homotopy Theory*, Springer-Verlag, 1967.

[35] Galligo A., Granger M., and Maisonobe P., *$\mathfrak{D}$-modules et faisceaux pervers dont le support singulier est un croissement normal*, preprint, Univ. de Nice (1983).

[36] Gelfand S. and MacPherson R., Verma Modules and Schubert Cells: A Dictionary, *Seminaire d'Algebra*, Lecture Notes in Math. **924**, Springer-Verlag, Berlin–Heidelberg–New York, 1982.

[37] Goresky M., Whitney Stratified Chains and Cochains, *Trans. Amer. Math. Soc.* **267** (1981), pp. 175–196.

[38] Goresky M., Intersection Homology Operations, to appear in *Comment. Math. Helv.*

[39] Goresky M. and MacPherson R., La dualité de Poincaré pour les espaces singuliers, *C. R. Acad. Sci. (Serie A)* **284** (1977), pp. 1549–1551.

[40] Goresky M. and MacPherson R., On the Topology of Complex Algebraic Maps,

Algebraic Geometry — Proceedings, La Rabida, Springer Lect. Notes in Math. **961** (1982), pp. 119–129.

[41] Goresky M. and MacPherson R., *Morse Theory and Intersection homology,* to appear in [6], above.

[42] Goresky M. and MacPherson R., *The Lefschetz Fixed Point Theorem and Intersection Homology,* to appear.

[43] Goresky M. and Siegel P., Linking Pairings on Singular Spaces, *Comment. Math. Helv.* **58** (1) (1983), pp. 96–110.

[44] Habegger N., *Obstructions to Immersions in Intersection Homology Theory,* preprint.

[45] Hardt R., Topological Properties of Subanalytic Sets, *Trans. Amer. Math. Soc.* **211** (1975), pp. 57–70.

[46] Hartshorne R., *Residues and Duality,* Springer Lecture Notes in Math. **20** (1966).

[47] Haskell P. *Index Theory on Curves,* to appear.

[48] Hirzebruch F., *New Topological Methods in Algebraic Geometry,* Springer-Verlag, 1966.

[49] Hsiang W. C. and Pati V., *L^2 Cohomology of Normal Algebraic Surfaces,* preprint.

[50] Kashiwara, *The Riemann–Hilbert Problem,* preprint, R.I.M.S. (1983).

[51] Kazhdan D. and Lusztig G., Representations of Coxeter Groups and Hecke Algebras, *Inv. Math.* **53** (1979), pp. 165–184.

[52] Kazhdan D. and Lusztig G., Schubert Varieties and Poincaré Duality, *Proc. Symp. Pure Math.* **36,** pp. 185–203, Amer. Math. Soc., 1980.

[53] Lefschetz S., *Topology,* Am. Math. Soc., 1930.

[54] Lusztig G., Green Polynomials and Singularities of Unipotent Classes, *Adv. in Math.* **42** (1981), pp. 169–178.

[55] Lusztig G., Some Problems in the Representation Theory of Finite Chevalley Groups, *Proc. Symp. Pure Math.* **37,** Amer. Math. Soc. (1980), pp. 313–317.

[56] Lusztig G., *Equivariant Sheaves on Reductive Groups I,* preprint.

[57] Lusztig G. and Vogan D., Singularities of Closures of K-Orbits on Flag Manifolds, *Inventiones* **71** (2) (1983), pp. 365–380.

[58] Łojasiewicz S., Sur les ensembles semi-analytiques, *Acts Int. Cong. Math.* (Nice, 1970) II, Gauthier-Villars, pp. 237–241.

[59] MacPherson R., Chern Classes for Singular Algebraic Varieties, *Annals of Math.* **100** (1974), pp. 423–432.

[60] MacPherson R. and Vilonen K., *Construction élémentaire des faisceaux pervers,* preprint, 1983.

[61] Malgrange B., Sur les points singuliers des équations differentielles, *Enseign. Math.* **XX** (1974), pp. 147–176.

[62] Mather J., *Stratifications and Mappings in Dynamical Systems* (edited by M. M. Peixoto), Academic Press, New York, 1973.

[63] Mebkhout Z., Thèse d'Etat, Université de Paris VII, février 1979.

[64] Morgan J., to appear.

[65] Pardon W., to appear.

[66] Pham F., *Singularité des systèmes différentiels de Gauss–Manin,* Progress in Mathematics, Birkhäuser (1979).

[67] Pignoni R., Density and Stability of Morse Functions on a Stratified Space, *Ann. Scuola Nor. Sup. Pisa (ser. IV)* **6** (1979), pp. 593–608.

[68] Poincaré H., Analysis Situs, *Journal de l'Ecole Polytechnique* **1** (1895), pp. 1–121. (Reprinted in *Oeuvres complètes,* pp. 193–288).

[69] SGA 7, *Groupes de monodromie en géometrie algébrique*, by A. Grothendieck, P. Deligne, and N. Katz, Lect. Notes in Math. **288** and **340**, Springer-Verlag, 1972, 1973.
[70] Siegel P., Witt Spaces, a Geometric Cycle Theory for KO Homology at Odd Primes, Ph.D. Thesis (M. I. T.) 1979 (to appear in *Amer. J. of Math.*).
[71] Slodowy P., Four Lectures on Simple Groups and Singularities, *Comm. Math. Inst. Utrecht* (1980).
[72] Springer T. A., Trigonometric Sums, Green Functions of Finite Groups and Representations of Weyl Groups, *Inv. Math.* **36** (1976), pp. 173–207.
[73] Sullivan D., Combinatorial Invariants of Analytic Spaces, *Proceedings of Liverpool Singularities Symposium I*, Lect. Notes in Math. **192**, Springer-Verlag, 1970, pp. 165–168.
[74] Thom R., Ensembles et morphismes stratifiés, *Bull. Amer. Math. Soc.* **75** (1969), pp. 240–284.
[75] Verdier J.-L., Le théorème de dualité de Poincaré, *C. R. Acad. Sci.* **256** (1963).
[76] Verdier J.-L., The Lefschetz Fixed Point Formula in étale Cohomology, *Proc. of Conference on Local Fields*, Springer-Verlag, 1967.
[77] Verdier J.-L., *Categories derivées, Etat* 0, *SGA* 4½, Lect. Notes in Math. **569**, Springer-Verlag, 1977.
[78] Verdier J.-L., Spécialisation de faisceaux et monodromie modérée, to appear in *Astérisque*.
[79] Vogan D., Irreducible characters of Semisimple Lie Groups III. Proof of the Kazhdan–Lusztig conjectures in the Integral Case, *Inv. Math.* **71** (2) (1983), pp. 381–418.
[80] Whitney H., Tangents to an Analytic Variety, *Ann. of Math.* **81** (1964), pp. 496–549.
[81] Zelevinsky A., A p-Adic Analogue of the Kazhdan–Lusztig Conjecture (in Russian), *Funct. Anal. and Its Appl.* **15** (1981), pp. 9–21.
[82] Zucker S., Hodge Theory Degenerating Coefficients I, *Ann. of Math.* **109** (1979), pp. 415–476.
[83] Zucker S., L_2-Cohomology of Warped Products and Arithmetic Groups, *Inventiones Math.* **70** (1982), pp. 169–218.
[84] Zucker S., *L_2-Cohomology and Intersection Homology of Locally Symmetric Varieties II*, preprint (1983).

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

ALEKSANDER PEŁCZYŃSKI

Structural Theory of Banach Spaces and Its Interplay with Analysis and Probability

I. Introduction

In this survey we consider Banach spaces as a branch of linear functional analysis (some other branches are: operator theory in Hilbert spaces and C^*-algebras, abstract harmonic analysis, Banach algebras, general theory of linear operators, linear topological spaces). The purpose of this survey is to give some idea what the modern theory of Banach spaces is about. We shall concentrate mainly on the structural problems of the theory, occasionally touching upon operator-theoretic aspects.

Recent progress in Banach spaces is due to introducing new techniques: function theoretic, probabilistic and combinatorial. In a few cases they have led also to new applications of Banach space techniques in these fields.

II. Historical remarks

The theory of Banach spaces is a legitimate child of the Polish Mathematical School in Lwów. The date of its birth is not exactly known. For bureaucratic reasons the appearance of the French edition of Banach's "Theorie des opérations lineaires" in 1932 is regarded as the date. The pioneering work of Polish mathematicians (S. Banach, H. Steinhaus, S. Mazur, W. Orlicz, J. Schauder) was abruptly stopped by the outbreak of the Second World War. However, at that moment the usefulness of the language of Banach spaces had been recognized in various branches of modern analysis: function theory, differential equations, harmonic analysis, approximation theory, summability theory. During the first 20 years after the war Banach spaces were treated like an old member of the academy, who deserves esteem for his merits but has nothing more to contribute. Attention was focussed on Banach algebras, locally convex spaces and abstract harmonic analysis.

The fundamental papers of A. Grothendieck published around 1955 marked the rebirth of the theory. They gradually came to be understood only in the late sixties. Among the mathematicians active already in the sixties who contributed to the renaissance of the theory one should mention first of all R. C. James, M. I. Kadec, J. Lindenstrauss, A. Pietsch and H. P. Rosenthal. Modern history starts with Enflo's solution of the approximation problem in 1972 (cf. Enflo, 1973) and with the Great French Revolution in Banach spaces.

III. Terminology and notation

Our notation is similar to that used in the monographs of Day (1973), Dunford and Schwartz (1958), Lindenstrauss and Tzafriri (1977), (1979). "*Space*" means either a real or a complex Banach space; "*n-space*" means "n-dimensional space" ($n = 1, 2, \ldots,$); "*operator*" means "bounded linear operator"; "*projection*" means "bounded linear projection"; "*isomorphism*" means a linear homeomorphism (onto!). $X^*, X^{**}, \ldots$ denote consecutive duals of a space X.

The *Banach–Mazur distance* $d(X, Y)$ between X and Y is

$$d(X, Y) = \inf\{\|T\|\|T^{-1}\| \,|\, T\colon X \to Y \text{ isomorphism}\};$$

we adopt the convention $\inf\emptyset = +\infty$.

"*X is isometric to Y*" means "$\exists T\colon X \to Y$ isomorphism with $\|T\|\,\|T^{-1}\| = 1$"; "*X is k-isomorphic to Y*" means "$d(X, Y) \leqslant k$ $(1 \leqslant k < +\infty)$". A subspace E of X is *k-complemented* in X if it is a range of a projection from X of norm $\leqslant k$. "*E is complemented in X*" means "E is k-complemented in X for some k".

Let $\boldsymbol{K}$ be an unspecified field of scalars, either $\boldsymbol{R}$*—reals* or $\boldsymbol{C}$*—complex numbers*. The finite-dimensional analogs of the classical sequence spaces l^p $(1 \leqslant p \leqslant \infty)$ and c_0 are the spaces $l_n^p = (\boldsymbol{K}^n, \|\ \|_p)$, where for $x = (t_j) \in \boldsymbol{K}^n$,

$$\|x\|_p = \Big(\sum_{j=1}^{n} |t_j|^p\Big)^{1/p} \quad \text{for} \quad 1 \leqslant p < \infty \quad \text{and} \quad \|x\|_\infty = \max_{1\leqslant j\leqslant n} |t_j|$$

$(n = 1, 2, \ldots)$. By an "*absolutely convex body*" we mean an "absolutely convex closed bounded neighbourhood of zero in $\boldsymbol{K}^n$".

Recall that a space X has the *approximation property* if the identity on X belongs to the closure of finite rank operators in a compact open topology. A sequence $(e_j) \subset X$ is a *basis* for X if for every $x \in X$ there is a unique sequence of scalaras (c_j) with $x = \sum_j c_j e_j$. If for every $x \in X$ the series converges unconditionally, the basis is called *unconditional*. Spaces with unconditional bases are isomorphic to *Banach lattices* (but not *vice versa*!). For definitions of real and complex Banach lattices cf. Lindenstrauss and Tzafriri (1979, Chap. 1). This book contains an excellent presentation of Banach lattices from the point of view of geometry and structural theory of Banach spaces. For additional information the reader may consult a survey by Bukhvalov, Veksler and Lozanovskii (1979), books by Lacey (1974) and Schaefer (1974), memoir by Johnson, Maurey, Schechtman and Tzafriri (1979) and recent papers by Pisier (1979), Bourgain and Talagrand (1981), Talagrand (1983).

IV. The frontier of Banach lattices

Banach lattices are by now the best understood, thoroughly investigated class of Banach spaces. Most of the function and sequence spaces have the natural structure of a Banach lattice. However, sometimes the existence of a lattice structure is not obvious at all. This is the case as regards the classical Hardy space $H^1(D) = H^1$ and its analogues in several variables. Recall that H^1 is the space of analytic functions in the open disc D such that

$$\sup\left\{(2\pi)^{-1}\int_0^{2\pi} |f(re^{it})|\,dt \,\middle|\, 0 < r < 1\right\} < +\infty.$$

Maurey (1980), using the so-called Banach space decomposition method, proved that H^1 is isomorphic to the space MartH^1 ($=$ Martingale H^1). The latter can be defined as the space of all functions f in $L^1[0,1]$ whose expansions $\sum c_j h_j$ with respect to the Haar system (h_j) converge unconditionally in L^1 with the norm

$$|||f||| = \sup\left\{\int_0^1 \left|\sum \varepsilon_j c_j h_j\right| dt \,\middle|\, \varepsilon_j = \pm 1;\ j = 1, 2, \ldots\right\}.$$

Clearly, the Haar system is an unconditional basis for MartH^1. Thus H^1 has an unconditional basis. However Maurey's original proof was practically non-effective and gave no information what an unconditional basis in H^1 looks like. An unconditional basis was constructed by Carleson

(1980). Later Wojtaszczyk (1982) proved that the classical Franklin orthonormal system forms an unconditional basis for H^1. Wojtaszczyk combined Carleson's method with the estimates of individual Franklin functions due to Ciesielski (1966) (cf. also Bochkarev, 1974). The coefficient functionals of the Franklin basis are in the space VMO (= functions with the vanishing mean oscilation) — a predual of H^1. Thus the spaces VMO, H^1 and BMO (= the dual of H^1) are isomorphic to Banach lattices.

Let D^n and B_n denote the unit polydisc and the unit ball in C^n respectively. Using the purely formal argument that the natural L^1-tensor product of two subspaces of L^1 with unconditional bases has an unconditional basis we conclude that the spaces $H^1(D^n)$ have unconditional bases for $n = 1, 2, \ldots$ (note that if $n \neq m$ then $H^1(D^n)$ is not isomorphic to $H^1(D^m)$ (Bourgain, 1983b)). Much deeper is the fact that $H^1(B_n)$ has an unconditional basis because $H^1(B_n)$ is isomorphic to H^1 for $n = 1, 2, \ldots$ (Wojtaszczyk, 1983). Again Wojtaszczyk's proof uses the decomposition method; it also uses along with some other analytic ingredients A. B. Aleksandrov's (1982) and Løw's (1982) result on the existence of non-trivial inner functions on B_n. However, we still do not know what the unconditional basis in $H^1(B_n)$ looks like. In particular, it is not known whether there is an orthonormal system in $H^2(B_n)$ which forms an unconditional basis for $H^1(B_n)$.

It seems interesting to decide whether the Sobolev space $W_{1,1}(T^2)$, consisting of periodic functions on 2-dimensional torus with the norm $\int_{T^2} (|f| + |D_x f| + |D_y f|)\, d\xi\, d\eta$ is isomorphic to a Banach lattice. In many aspects $W_{1,1}(T^2)$ behaves like H^1. However, $W_{1,1}(T^2)$ has no unconditional basis because it contains a subspace isomorphic to L^1.

The following open problem is well known:

Is a complemented subspace of a Banach lattice isomorphic to a Banach lattice?

The answer is unknown even for Banach spaces with unconditional bases and for complemented subspaces of $L^p[0, 1]$ for $1 < p \neq 2 < \infty$. All known complemented subspaces of $L^p[0, 1]$ have unconditional bases. For every fixed p as above there are at least uncountably many mutually non-isomorphic subspaces of this kind (Bourgain, Rosenthal and Schechtman, 1981; cf. also Bourgain, 1981). The proof of the latter result is delicate: it uses classical martingale inequalities due to Burkholder, Davis and Gundi, and Stein combined with techniques borrowed from descriptive set theory.

V. Important classes of Banach spaces without a lattice structure

There are important Banach spaces with a rich additional structure which are not Banach lattices. The next result exhibits some of them.

THEOREM. *The following Banach spaces and their duals are not isomorphic to complemented subspaces of Banach lattices*:

(a) $B(l^2)$, *and* $K(l^2)$ — *the spaces of all bounded, respectively all compact operators on* l^2 (Gordon and Lewis, 1974);

(b) *the disc algebra* A, *and* H^∞ — *the spaces of uniformly continuous, respectively bounded analytic functions in the open unit disc* (Pełczyński, 1974);

(c) $C^{(k)}(T^n)$ — *the space of all k-times differentiable periodic functions of n-variables*, $k \geqslant 1$, $n \geqslant 2$ (Kisliakov, 1977; Kwapień and Pełczyński, 1980).

The proof is based on the fact, discovered by Gordon and Lewis (1974), that 1-summing operators (cf. Section XII) from a Banach lattice factor through $L^1(\mu)$-spaces. To prove the lack of this property in cases (a), (b), (c) one has to use some special analytic properties of the spaces under consideration. For instance for (c) one uses the boundedness of the Sobolev embedding in a limit case and the fact that the orthogonal projection from vector-valued $L^2(T^n)$ onto the natural representation of the Sobolev space $W_{k,2}(T^n)$ in that space is of weak type (1, 1).

The theorem can be generalized in various directions. Case (a) extends to C^*-algebras which contain for all n uniformly complemented subspaces uniformly isomorphic to the space $B(l_n^2)$ of all bounded operators on l_n^2; it also generalizes to various spaces of operators, unitary ideals and translation invariant spaces (cf. Lewis, 1975; Pisier, 1978; Kwapień and Pełczyński, 1980; Pełczyński and Schütt, 1981). Case (b) extends onto a wide class of uniform algebras, for instance all uniform algebras with at least one non-trivial Gleason part; the following seems plausible:

GENERAL GLICKSBERG CONJECTURE. *A uniform algebra is isomorphic to a Banach lattice iff it is self-adjoint* (cf. Pełczyński, 1977).

Case (c) extends to spaces of k-times continuously differentiable functions on n-dimensional differentiable manifolds ($k \geqslant 1$, $n \geqslant 2$). One applies a result of Mitjagin (1970) on the isomorphism of such spaces.

VI. Banach spaces containing l^p

The simplest infinite-dimensional Banach spaces are l^p ($1 \leqslant p < \infty$) and c_0 spaces. Therefore it is natural to ask whether every infinite-dimensional Banach space contains some of them. (We adopt the convention that "X contains Y" means "X contains a subspace isomorphic to Y".) The answer is in general negative; as we shall see later (cf. Section X), the situation is entirely different in the case of l_n^p spaces.

We have

THEOREM (Tsirelson, 1974). *There exists an infinite-dimensional Banach space X which contains no l^p ($1 \leqslant p < \infty$) and no c_0; X is reflexive and has an unconditional basis.*

Figiel and Johnson (1974), inspired by Tsirelson, constructed similar uniformly convex spaces.

We refer the reader to Schechtman (1979), Casazza, Johnson and Tzafriri (1984) and the survey by Casazza [a] for further information.

Can one find a similar example among subspaces of infinite-dimensional $L^r(\nu)$-spaces ($1 \leqslant r \leqslant \infty$) and $C(S)$-spaces? This question — a fundamental structural question of classical Banach spaces — has a long history and is much older than the Tsirelson example. The last step towards a solution is due to Aldous (1981). The answer is contained in the table ($L^r = L^r[0, 1]$; "subspace" = infinite-dimensional subspace; here "contains l^r" has a strong meaning: "$\forall \varepsilon > 0$ $\exists$ subspace Y with $d(Y, l^r) < 1+\varepsilon$").

The result of Bretagnolle, Dacunha-Castelle and Krivine (1966) is based upon P. Levy's theory of p-stable random variables. The result of Aldous (1981) uses probabilistic ideas about domains of attraction combined with an ingenious compactness argument; it is also based upon the results from a remarkable paper by Rosenthal (1973).

Analysing Aldous' argument, Krivine and Maurey (1981) gave an alternative proof of his result: they introduced the concept of "stable spaces" defined in terms of commutativity of certain iterated limits with respect to ultrafilters. Every stable space contains l^p for some $p \in [1, \infty)$; the class of stable spaces is hereditary with respect to subspaces; L^r are stable for $1 \leqslant r < \infty$. Using the technique of stable spaces, S. Guerre and M. Levy (1983) improved the result of Aldous. They showed that if X is an infinite-dimensional subspace of L^1 then, for every $\varepsilon > 0$, X contains a subspace Y with $d(Y, l^{p(X)}) < 1+\varepsilon$ where $p(X)$ is the type index of X (cf. Section X for the definition).

Classical Banach space	Answer	References
L^∞	L^∞ contains isometrically all separable Banach spaces	Banach and Mazur (1933)
$C(S)$, S-uncountable compact metric	$C(S)$ contains isometrically all separable Banach spaces	Banach and Mazur, cf. Banach (1932)
$C(S)$, S-countable compact	every subspace of $C(S)$ contains c_0	Pełczyński and Semadeni (1959)
$L^r, 2 < r < \infty$	every subspace either contains l^r or is isomorphic to l^2	Kadec and Pełczyński (1962)
	l^2 is contained in L^r	already known about 1920
L^2	every subspace is isometric to l^2	
$L^r, 1 \leqslant r < 2$	L^p is isometric to a subspace of L^r for $p \in [r, 2]$	essentially P. Levy (1937); Bretagnoll, Dacunha-Castelle and Krivine (1966).
	every subspace of L^r contains l^p for some $p \in [r, 2]$	Aldous (1981)
L^1	every non reflexive subspace contains l^1	Kadec and Pełczyński (1962)
	every reflexive subspace of L^1 is isomorphic to a subspace of L^p for some $p \in [1, 2]$	Rosenthal (1973)

VII. Infinite-dimensional subspaces of general Banach space; the main problems

Say that a Banach space X satisfies *James trichotomy*, in symbol $X \in \mathrm{JT}$, if X contains either l^1 or c_0 or an infinite-dimensional reflexive subspace; X has the *unconditional basic sequence property*, in symbol $X \in \mathrm{UBSP}$, if X has an infinite-dimensional subspace with an unconditional basis. The fundamental problems about infinite-dimensional subspaces are:

P.1. *Let X be an infinite-dimensional Banach space. Do we have $X \in \mathrm{JT}$?*

P.2. *Let X be an infinite-dimensional Banach space. Do we have $X \in \mathrm{UBSP}$?* (Mazur, about 1955, unpublished).

By a classical result due to James (1950), $X \in \mathrm{UBSP} \Rightarrow X \in \mathrm{JT}$. $X \in \mathrm{UBSP}$ provided that either (a) X is a subspace of a space with an unconditional basis (Bessaga and Pełczyński, 1958), or (b) X is a subspace of a "nice" Banach lattice, e.g., an order-continuous lattice (Figiel, Johnson and Tzafriri, 1975). Moreover, $X \in \mathrm{JT}$ provided X^{**} is separable (Johnson and Rosenthal, 1972). An example of Maurey and Rosenthal (1977) of a normalized weakly null sequence which does not have unconditional basic subsequences indicates that the answer to P.2 might be negative.

In connection with P.1 we note that there are many examples, starting from James' quasi-reflexive space (cf. James, 1951), of non-reflexive spaces which do not contain l^1 or c_0. Among the more recent ones is the example of Bourgain and Delbaen (1980) of a space which does not contain c_0 but whose dual is isomorphic to l^1.

Essential progress towards the solution of P.1 was made thanks to the Rosenthal dichotomy criterion:

THEOREM (Rosenthal, 1974). *Every bounded sequence in a Banach space either contains a subsequence equivalent to the unit vector basic of l^1 or contains a weak Cauchy subsequence.*

Originally Rosenthal established the result for real spaces. An additional trick to extend his proof to the complex case was given by Dor (1975). Rosenthal's proof uses a technique of infinite combinatorics. Farahat (1974) simplified the proof by using the so-called Ramsey property of analytic subsets of 2^N.

For a further generalization of Rosenthal's theorem we refer to Odell and Rosenthal (1975), Rosenthal (1977), Bourgain, Fremlin and Talagrand (1978); the last paper gives a delicate result on compactness, with respect to pointwise convergence of sets of functions of the first Baire class. We refer to the excellent surveys by Odell (1980) and Maurey (1983) for additional information.

VIII. Other infinite-dimensional problems

A. Isomorphic classification of complemented subspaces of $L^p = L^p[0,1]$ $(1 \leqslant p \leqslant \infty)$ and $C = C[0,1]$. The case $1 < p < \infty$ has already been mentioned at the end of Section IV. All known complemented subspaces of L^1 are either isomorphic to L^1 or isomorphic to l^1 or finite-dimensional; if X is complemented in L^1 and X has the Radon–Nikodym property (in particular, if X is isomorphic to a subspace of a separable dual) then either X is isomorphic to l^1 or X is finite-dimensional (Lewis and Stegall, 1973). Every complemented subspace of L^∞ is either isomorphic to L^∞ or finite-dimensional (Lindenstrauss, 1967). All known complemented subspaces of C are isomorphic to $C(S)$-spaces for S compact and metric; if X^* is complemented in C and X is not separable, then X is isomorphic to C (Rosenthal, 1972). There are exactly ω_1 mutually non-isomorphic $C(S)$-spaces for S compact and metric (Milutin, 1966; Bessaga and Pełczyński 1960), whereas there are exactly two different types of isomorphisms of spaces of all continuous scalar-valued bounded functions on a non-compact separable complete metric space (Etcheberry, 1975; Khmieleva, 1981). For an infinite-dimensional separable Banach space X the following properties are equivalent: (i) X is isomorphic to c_0, (ii) every subspace of C isomorphic to X is complemented in C, (iii) every subspace isomorphic to X of any separable Banach space Y is complemented in Y; the implications (i)$\Rightarrow$(ii) and (i)$\Rightarrow$(iii) are due to Sobczyk (1941), the others to Zippin (1977); cf. also Alspach (1978) and Benyamini (1978).

B. ***Find a complemented subspace of a Banach space with a basis which fails to have a basis.***

This problem is parallel to that of complemented subspaces of a Banach lattice mentioned in Section IV. Note that a Banach space is a complemented subspace of a Banach space with a basis iff X has BAP, i.e. if there is a sequence $(T_n\colon X \to X)$ of finite rank operators with $\lim_n \|T_n(x) - x\| = 0$ for all $x \in X$ (Pełczyński, 1971; Johnson, Rosenthal and Zippin, 1971).

This problem is also related to the following question in harmonic analysis. Let Ω be a subset of all integers. Let C_Ω denote the subspace of $C(\boldsymbol{T})$ generated by the characters $\{e^{int}\colon n \in \Omega\}$. Has C_Ω a basis? By the Féjer theorem, C_Ω has BAP.

C. The hyperplane problem of Banach. *Is every infinite-dimensional Banach space isomorphic to its subspace of codimension one?*

Johnson, Lindenstrauss and Schechtman (1980) conjecture that the space constructed by Kalton and Peck (1979) provides a counterexample.

D. The distorted norm problem. *Let X be isomorphic to l^p $(1 < p < \infty)$. Does X contain for every $\varepsilon > 0$ a subspace Y_ε with $d(Y_\varepsilon, l^p) < 1 + \varepsilon$?*

For more information concerning D we refer to Lindenstrauss and Tzafriri (1973), (1977), and Maurey (1983).

E. The three space problem and the semi-embedding problem for L^1.

(a) *Let X be a subspace of L^1. Is it true that either X or L^1/X contain a subspace isomorphic to L^1?*

(b) *Assume that there exists a one-to-one operator from L^1 into a Banach space Y which takes the unit ball of L^1 into a closed subset of Y. Does Y contain a subspace isomorphic to L^1?*

Problems (a) and (b) are closely related. For more information we refer to Bourgain and Rosenthal (1983), Bourgain (1983), Rosenthal (1983), [a], Ghoussoub and Rosenthal [a].

IX. A few introductory words on the local theory

Starting from this section until the end of our survey we deal with the so-called local theory. This is at present the most vigorous area of activity in Banach spaces.

The local theory of Banach spaces is concerned with the asymptotic behaviour of properties of finite-dimensional spaces and operators acting between them; asymptotic as the dimension of the spaces tends to infinity. In particular the theory deals with those isomorphic invariants (of infinite-dimensional spaces) which can be expressed in that way.

We shall try to illustrate what the local theory is about on a few selected examples. For comprehensive discussions of various aspects of local theory we refer to surveys by Maurey (1983) and Figiel (1984), and reports by Pisier (1982), (1984). General references for operator aspects of the theory are the seminar reports Sém. (1972/73), (1973/74), (1974/75), (1975/76), (1977/78), (1978/79), (1979/80), (1980/81) and the books by Pietsch (1978), (1982); special topics are presented in the lecture notes of Pełczyński (1980).

X. Type, cotype, and almost isometric l_n^p subspaces of a Banach space

One of the central problems of the local theory is to determine which infinite-dimensional spaces contain, for a given $p \in [1, \infty]$, all l_n^p.

To present results it is convenient to adopt the following slang:

"*X contains l_n^p uniformly*" means "for fixed $p \in [1, \infty]$ there exists $C < \infty$ such that for every n there is an n-space $E \subset X$ with $d(E, l_n^p) < C$";

"*X contains l_n^p almost isometrically*" means "for fixed $p \in [1, \infty]$ for every $\varepsilon > 0$ and for every n there is an n-space $E \subset X$ with $d(E, l_n^p) < 1 + \varepsilon$.

Now we are ready to state the following fundamental

THEOREM (Dvoretzky, 1959, 1961). *Every infinite-dimensional Banach space contains l_n^2 almost isometrically.*

Another important result is

THEOREM (Krivine, 1976). *An infinite-dimensional Banach space which contains l_n^p uniformly contains l_n^p almost isometrically.*

We shall return to Dvoretzky's theorem in Section XV. Concerning Krivine's theorem we note that the special cases $p = 1$ and $p = \infty$ were discovered by James (1964) and Giesy (1966), the proof in those cases is elementary. For $1 < p \neq 2 < \infty$ Krivine's proof is difficult, combinatorial in nature; some techniques borrowed from model theory are used. An alternative proof is due to Rosenthal (1978).

How to determine that given Banach space X contains, for a given $p \neq 2$, l_n^p almost isometrically? To give the answer we recall the important concepts of type and cotype of Banach spaces introduced in Hoffman-Jørgensen (1972), (1974).

DEFINITION. Let $1 \leqslant p \leqslant 2 \leqslant q \leqslant \infty$. A Banach space X is said to be *of type p* (resp. *of cotype q*) if there exists a constant t_p (resp. c_q) such that for all sequences $(x_j) \subset X$

$$\Big(\operatorname*{Average}_{\varepsilon_j=\pm 1}\Big\|\sum_j \varepsilon_j x_j\Big\|^2\Big)^{1/2} \leqslant t_p\Big(\sum_j \|x_j\|^p\Big)^{1/p} \tag{1}$$

resp.

$$\Big(\operatorname*{Average}_{\varepsilon_j=\pm 1}\Big\|\sum_j \varepsilon_j x_j\Big\|^2\Big)^{1/2} \geqslant c_q^{-1}\Big(\sum_j \|x_j\|^q\Big)^{1/q}. \tag{2}$$

The numbers

$$\inf\{t_p \mid \text{inequality (1) holds for all } (x_j) \in X\},$$

$$\inf\{c_q \mid \text{inequality (2) holds for all } (x_j) \in X\}$$

are called the *type p* and the *cotype q constants of X* respectively.

It is not hard to prove that every Banach space is of some type p with $1 \leqslant p \leqslant 2$ and of some cotype q with $2 \leqslant q \leqslant \infty$.

Following Maurey and Pisier (1976), we next introduce the *type* and the *cotype indices* of X,

$$p(X) = \inf\{p\colon X \text{ is of type } p\},$$

$$q(X) = \sup\{q\colon X \text{ is cotype } q\}.$$

(Note that a space X need not be of type $p(X)$ or cotype $q(X)$.)

Now we are ready to state a result which gives a satisfactory answer to our question.

THEOREM (Maurey and Pisier, 1976). *Every infinite-dimensional Banach space X contains almost isometrically $l_n^{p(X)}$ and $l_n^{q(X)}$ and X does not contain uniformly l_n^q for $q > q(X)$ and l_n^p for $p < p(X)$.*

The indices $p(X)$ and $q(X)$ measure in some sense how much X resembles a Hilbert space. First we have

THEOREM (Kwapień, 1972). *A Banach space is of type 2 and of cotype 2 iff it is isomorphic to a Hilbert space.*

Next we note the following:

Let $1 \leqslant r < \infty$. Then $p(L^r) = \min(2, r)$, $q(L^r) = \max(2, r)$; moreover, L^r is of type $p(L^r)$ and of cotype $q(L^r)$ (essentially Orlicz, 1933).

We end this section by mentioning a few remarkable examples of Banach spaces related to the concepts of type and cotype.

1. There is a non-reflexive space of type 2 and finite cotype (James, 1978).

2. If X is a Banach space such that every subspace of X has the approximation property, then $p(X) = q(X) = 2$ (essentially Szankowski, 1976, 1978). On the other hand, there is a space X of type 2 and cotype q for all $q > 2$ such that every subspace of every quotient of X has a basis but l^2 is not isomorphic to a subspace of a quotient of X (Johnson, 1980).

3. If X is a Banach space such that there is an exact sequence $0 \to E \to X \to X/E \to 0$ with E and X/E isomorphic to Hilbert spaces, then $p(X) = q(X) = 2$ (Enflo, Lindenstrauss and Pisier, 1975).

On the other hand there is a space X non-isomorphic to any Hilbert space which has a subspace E such that E and X/E are isomorphic to l^2 (Enflo, Lindenstrauss and Pisier, 1975); for another example cf. Kalton and Peck (1979).

We refer the reader to the book by Lindenstrauss and Tzafriri (1979, Chap. 1, e, f, g) for additional information.

The next two sections are devoted to the "historical roots" of type and cotype.

XI. Vector random series and probability in Banach spaces

Hoffman-Jørgensen (1972) originally introduced the concepts of type and cotype of a Banach space to study random series in Banach spaces. Note that

$$\operatorname*{Average}_{\varepsilon_j = \pm 1} \Big\| \sum_j \varepsilon_j x_j \Big\|^2 = \boldsymbol{E} \Big\| \sum_j \delta_j x_j \Big\|^2,$$

where E denotes the mathematical expectation and (δ_j) is a sequence of ± 1-valued mutually independent random variables each of mean zero. The condition that a Banach space X is of type p is equivalent to the statement that if $(x_j) \subset X$ and $\sum \|x_j\|^p < +\infty$ then the random series $\sum_j \delta_j x_j$ converges almost surely. Similarly one can restate that X is of cotype q. Now one replaces (δ_j) by another sequence of mutually independent equally distributed random variables (e.g. normal Gaussian variables, p-stable variables, etc.) and investigates the induced random series and corresponding concepts of type and cotype. This program has been implemented by Hoffman-Jørgensen (1974) and Maurey and Pisier (1976).

There is a natural question how to extend onto Banach space valued random variables the classical limit theorems like the central limit theorem, the law of large numbers and the law of the iterated logarithm. Again the "right answer" is given in terms of type and cotype. For example we have

THEOREM (Hoffman-Jørgensen and Pisier, 1976). *A Banach space X is of type 2 iff the central limit theorem holds for all sequences of identically distributed X-valued independent random variables of mean zero with a finite second moment.*

The first result of this nature goes back to Beck (1962). For further information we refer the reader to surveys by Hoffman-Jørgensen (1977), Woyczyński (1978) with a comprehensive literature until 1977, the memoir by Marcus and Pisier (1981) and the book by Arajuo and Gine (1980).

XII. Absolutely summing and hilbertian operators

The following important concept is due to Pietsch (1967):

DEFINITION. An operator $u: X \to Y$ is *r-summing* $(0 < r < \infty)$ if for some $C > 0$ and all sequences $(x_j) \subset X$

$$\sum \|u(x_j)\|^r \leqslant C^r \sup\Big\{\sum_j x^*(x_j)|^r : x^* \in X^*,\ \|x^*\| \leqslant 1\Big\}.$$

The property of r-summability of an operator is a rather restrictive one; for instance: if X and Y are Hilbert spaces then $u: X \to Y$ is r-summing for some r iff it is r-summing for all r and iff it is of the Hilbert–Schmidt type (Pełczyński, 1967); the composition of at least $[r]+2$ operators which are r-summing is a nuclear and therefore compact operator (Pietsch, 1967). The following result is important both for the theory of absolutely summing operators and for many applications, including Sobolev spaces and Hardy spaces.

THEOREM (Pietsch, 1967). *Let X be a subspace of a $C(S)$-space and Y an arbitrary Banach space and let $r \geqslant 1$. Then an operator $u\colon X \to Y$ is r-summing iff it is bounded with respect to some $L^r(\mu)$-norm — precisely, iff there is a probability Borel measure μ on S such that for some $C > 0$*

$$\|u(x)\|^r \leqslant C^r \int_S |x(s)|^r d\mu \quad \textit{for all } x \in X.$$

The next result links r-summing operators with the type and cotype:

THEOREM (Maurey, 1974). *Every bounded operator from a $C(S)$-space into a Banach space X of cotype q is r-summing for all $r > q$; moreover, it is 2-summing if $q = 2$. The order of summability is sharp.*

Special cases of Maurey's theorem $X = L^1$ and $X = L^2$ go back to Grothendieck (1956) (the case $X = L^1$ is equivalent to the famous Grothendieck inequality). It is not hard to extend Grothendieck's result to the case $X = L^p(\nu)$ for $1 < p < 2$ (cf. Lindenstrauss and Pełczyński, 1968). Next, Schwartz (1969) and Kwapień (1970) discovered the important special case of Maurey's theorem for $X = L^p(\nu)$, $p > 2$. Maurey's work was directly inspired by Rosenthal (1973).

Kislyakov (1976) and independently Pisier (1978a) extended Maurey's theorem by replacing a $C(S)$-space by its arbitrary subspace with a reflexive annihilator in $[C(S)]^*$; Bourgain (1983a) (cf. also Bourgain [a]) showed that the assertion of Maurey's theorem remains valid if the $C(S)$-space is replaced by the disc algebra. This is in fact a deep result in function theory.

For further information we refer to Maurey (1974), (1975), Krivine (1974), (1979), Pisier (1978b), Rosenthal (1976), the books by Lindenstrauss and Tzafriri (1977), (1979), Pietsch (1978).

Next we pass to hilbertian operators introduced by Grothendieck (1956).

DEFINITION. An operator $u\colon X \to Y$ is *hilbertian* if it admits a linear factorization

$$u\colon X \xrightarrow{v} H \xrightarrow{w} Y$$

through some Hilbert space H.

Working with hilbertian operators, one can use additional tools of the operator theory of Hilbert spaces. For instance let $Y = X$. Consider the operator

$$\tilde{u}\colon H \xrightarrow{w} X \xrightarrow{v} H$$

defined on the Hilbert space H. It appears that $\tilde{u}$ shares some properties with u, in particular u and $\tilde{u}$ have the same spectrum. The importance of this simple but useful trick was recognized by Pietsch (1963). He also observed that every 2-summing operator, and therefore every r-summing operator for $0 < r < 2$, is hilbertian.

These simple facts combined with the composition formula for various classes of operators and the Pietsch theory of s-numbers (cf. Pietsch, 1979, 1982; the first abstract scheme goes back to Mitjagin and Pełczyński, 1967) have been used to obtain various deep results concerning eigenvalues of p-summing operators and its relatives, abstract analogs of the classical Weyl and Horn inequalities and important new composition formulae (cf. Johnson, König, Maurey and Retherford, 1979; König, Retherford and Tomczak-Jaegermann, 1980; Pietsch, 1980, 1982; Kaiser and Retherford, 1983).

Again the type and cotype turn out to be useful for hilbertian operators.

THEOREM (Maurey, 1974). *Every operator from a space of type 2 to a space of cotype 2 is hilbertian.*

This result generalizes both Kwapień's theorem (cf. Section X) and the fact that if $\infty > p \geqslant 2 \geqslant q \geqslant 1$ then every operator from $L^p(\mu)$ to $L^q(\nu)$ is hilbertian (Lindenstrauss and Pełczyński, 1968).

A natural and important generalization of Hilbertian operators are operators which admit factorizations through an $L^p(\mu)$-space for fixed $p \in [1, \infty)$ (cf. Kwapień, 1972a; Krivine, 1974; Maurey, 1974; Lindenstrauss and Tzafriri, 1979; Pietsch, 1979, 1982).

Absolutely summing operators and their relatives have proved to be useful in several fields of analysis. Here are some references: Pisier (1978c), Kwapień and Pełczyński (1980) (harmonic analysis), Bourgain (1983a), [a], Pełczyński (1977) (spaces of analytic functions and uniform algebras), Varopoulos (1974), Ringrose (1975), Tonge (1976), Bennett (1977), Peller (1983) , Haagerup (1983) (operator algebras, C^*-algebras), König (1977), Pietsch (1979) (Sobolev embeddings), Ørno (1976) (real variables).

XIII. Recent results on hilbertian operators

The Grothendieck inequality mentioned in Section XII yields the following

THEOREM (Grothendieck, 1956). *Every operator from a $C(S)$-space into an $L^1(\mu)$ space is hilbertian.*

A far-reaching common generalization of this result and Maurey's theorems mentioned in Section XII is

THEOREM (Pisier, 1980). *If X^* and Y are of cotype 2 then every finite rank operator $u: X \to Y$ admits a factorization $u: X \xrightarrow{v} H \xrightarrow{w} Y$ with $\|v\|\,\|w\| \leqslant C$ where the constant C depends only on the cotype 2 constants of X^** and *Y.*

COROLLARY. *If in addition either X or Y satisfies the approximation property then every operator from X into Y is hilbertian.*

The main ingredient of Pisier's proof is the technique, which he developped, of vector-valued Riesz products on dyadic groups. It is combined with the extrapolation technique used in the proof of Maurey's theorem (the first mentioned in Section XII).

Comments. 1° Example (Pisier, 1983a; cf. also Pisier, 1984). *There is a Banach space X such that both X and X^* are of cotype 2 but the identity operator,* id: *$X \to X$, is not hilbertian; equivalently X is not isomorphic to a Hilbert space.*

This example shows that in Pisier's theorem some assumptions on the approximation property are in general necessary.

2° $B(l^2)$ fails the approximation property (Szankowski, 1981); however, every operator $u: B(H) \to [B(H)]^*$ is hilbertian (Haagerup, 1981). Note that $[B(H)]^*$ is of cotype 2 (Tomczak-Jaegermann, 1972) hence the main assumption of Pisier's theorem is fulfilled.

3° It is not known whether H^∞ has the approximation property. However, every operator $u: (H^\infty) \to (H^\infty)^*$ is hilbertian (Bourgain [a]). The deep result that $(H^\infty)^*$ is of cotype 2 is due to Bourgain (1983a).

4° Note that if X is of type 2 then X^* is of cotype 2, and compare Pisier's theorem with Maurey's theorem (the second one in Section XII), where no approximation property assumption is needed.

We end this section with open problems.

(a) *What reasonable additional assumptions are needed to extend Pisier's theorem to all operators?*

(b) *Is every operator u: $C^1(\boldsymbol{T}^2) \to [C^1(\boldsymbol{T}^2)]^*$ hilbertian?*

XIV. Local theory and classical operators

We begin with some examples:

(i) *The Fourier Transform $\mathscr{F}$: $L^2(\boldsymbol{R}) \to L^2(\boldsymbol{R})$,*

$$\mathscr{F}(f) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} f(t) e^{-its} dt. \tag{1}$$

Now we are given a Banach space X. For a "nice" X-valued test function $f \in L^2(\boldsymbol{R}, X)$ we define $\mathscr{F}_X(f)$ by (1). Here the symbol $L^2(\boldsymbol{R}, X)$ denotes the space of all Bochner square-integrable X-valued functions on $\boldsymbol{R}$. It is natural to ask for which Banach spaces X one obtains in this way a bounded linear operator $\mathscr{F}_X\colon L^2(\boldsymbol{R}, X) \to L^2(\boldsymbol{R}, X)$.

The answer is provided by the following

THEOREM (Kwapień, 1972). *$\mathscr{F}_x$ is a bounded operator iff X is isomorphic to a Hilbert space.*

(ii) *The Paley projection* $\mathscr{G}\colon L^2(\boldsymbol{T}) \to L^2(\boldsymbol{T})$,

$$\mathscr{G}(f)(s) = \sum_{j=1}^{\infty} \hat{f}(2^j) e^{i2^j s} \quad \text{for } f \in L^2(\boldsymbol{T}), \tag{2}$$

where

$$\hat{f}(k) = (2\pi)^{-1} \int_0^{2\pi} f(t) e^{-ikt} dt \quad (k = 0, \pm 1, \pm 2, \ldots).$$

Given a Banach space X and a "nice" function $f \in L^2(\boldsymbol{T}, X)$, for instance f being a trigonometric polynomial

$$f = \sum_{j=-n}^{n} x_j e^{-ijt}, \quad x_j \in X,\ j = 0, \pm 1, \pm 2, \ldots, \pm n,\ n = 0, 1, 2, \ldots$$

we define $\mathscr{G}_X(f) \in L^2(\boldsymbol{T}; X)$ by (2). Again we ask for which Banach spaces the operator defined in this way on a dense subspace of the space $L^2(\boldsymbol{T}; X)$ of all Bochner square-integrable X-valued functions extends to a bounded operator $\mathscr{G}_X\colon L^2(\boldsymbol{T}; X) \to L^2(\boldsymbol{T}; X)$. The answer is far from being obvious; it is given by the following

THEOREM (Pisier, 1982a). *$\mathscr{G}_X$ is bounded iff X has a type > 1.*

In other words, by the Maurey–Pisier theorem (cf. Section XI), $\mathscr{G}_X$ is bounded if X does not contain uniformly l_n^1. Combining Pisier's theorem with a result due to Figiel and Tomczak-Jaegermann (1979), we obtain the following

COROLLARY. *$\mathscr{G}_X$ is bounded iff X is π-Euclidean.*

Recall that a Banach space X is said to be *π-Euclidean* if there is a $\lambda > 0$ such that for every $k = 1, 2, \ldots$ there is a $\varphi(k) > 0$ such that every n-space $E \subset X$ with $n > \varphi(k)$ contains a k-dimensional subspace which is 2-isomorphic to l_k^2 and λ-complemented in X (Pełczyński and Rosenthal, 1975).

(iii) *The Hilbert transform* $\mathscr{H}$: $L^2(\boldsymbol{R}) \to L^2(\boldsymbol{R})$,

$$\mathscr{H}(f)(s) = p.v.\pi^{-1} \int\limits_{-\infty}^{+\infty} f(s-t)t^{-1}dt = \lim_{\varepsilon\to 0} \pi^{-1} \int\limits_{|t|>\varepsilon} f(s-t)t^{-1}dt. \tag{3}$$

Given a Banach space X and a "nice" test function $f \in L^2(\boldsymbol{R}; X)$, one can define $\mathscr{H}_X(f) \in L^2(\boldsymbol{R}; X)$ by (3). For which Banach spaces X does the operator defined on a dense set of "nice" test functions extend to a bounded linear operator $\mathscr{H}_X$: $L^2(\boldsymbol{R}; X) \to L^2(\boldsymbol{R}; X)$? The answer is of a different nature from those in the cases previously discussed. We shall need the following

DEFINITION (Burkholder, 1966). Given a Banach space X, we write $X \in \text{UMD}$ if for every $p \in (1, \infty)$ there exists a constant $c_p > 0$ such that for all $(x_j) \subset X$

$$\sup_{\varepsilon_j=\pm 1} \int\limits_0^1 \Big\|\sum_j \varepsilon_j x_j h_j(t)\Big\|^p dt \leqslant c_p \int\limits_0^1 \Big\|\sum_\sigma x_j h_j(t)\Big\|^p dt, \tag{3a}$$

(h_j) denotes the Haar orthonormal system.

In other words, $X \in \text{UMD}$ if the decomposition of $L^2([0,1]; X)$ with respect to the Haar system is unconditional. It is not hard to show that if $X \in \text{UMD}$ then inequality (3a) holds for all martingale difference sequences (d_j) (i.e., we replace in (3a) (h_j) by (d_j)). Thus we can say that X has the unconditional martingale difference property, which justifies the notation "$X \in \text{UMD}$". The following result answers our question:

THEOREM. *For every Banach space X the following conditions are equivalent:*

(α) $\mathscr{H}_X$: $L^2(\boldsymbol{R}; X) \to L^2(\boldsymbol{R}; X)$ *is bounded,*

(β) $X \in \text{UMD}$,

(γ) *there exists a symmetric biconvex function ξ on $X \times X$ satisfying $\xi(0,0) > 0$ and $\xi(x,y) \leqslant \|x+y\|$ whenever $\|x\| \leqslant 1 \leqslant \|y\|$.*

The equivalence (β)$\Leftrightarrow$(γ) is due to Burkholder (1981); the implication (β)$\to$(α) is due to Burkholder (1982) and (α)$\Rightarrow$(β) is due to Bourgain (1983c). The proofs involve both probabilistic and analytic techniques dealing with singular operators. Note that the proof of the implication (α)$\Rightarrow$(β) (for $X = \boldsymbol{R}$) provides a new argument for the classical result of Paley and Marcinkiewicz on the unconditionality of the Haar system in $L^p[0,1]$. The Hilbert transform in (α) can be replaced by many other singular operators (for instance by the Hilbert transform on the circle). The description of this class of operators is not clear at present. For additional information we refer the reader to Aldous (1979), Bourgain [b], Maurey (1975a), Pisier (1975).

The examples discussed above suggest the following general scheme. Given an operator $\mathscr{A}: L^2(\mu)\to L^2(\mu)$ and a Banach space X. Denote by $\mathscr{A}\otimes \mathrm{id}_X$ the operator which acts on finite combinations $z = \sum_{j=1}^{n} f_j\cdot x_j$, $f_j\in L^2(\mu)$, $x_j\in X$ $(j = 1, 2, \ldots, n,\ n = 1, 2, \ldots)$ as follows: $\mathscr{A}\otimes \mathrm{id}_X = \sum_{j=1}^{n} \mathscr{A}(f_j)x_j$.

Let $B_{\mathscr{A}}$ be the class of all Banach spaces X such that the operator $\mathscr{A}\otimes\mathrm{id}_X$ extends to a bounded linear operator $\mathscr{A}_X$: $L^2(\mu; X)\to L^2(\mu; X)$ ($L^2(\mu; X)$ denotes the space of Bochner square integrable X-valued functions).

The examples indicate that the investigation of $B_{\mathscr{A}}$ may lead to a better understanding of the geometry of Banach spaces as well as of classical operators.

A slightly different scheme is considered in Figiel (1984).

XV. Dvoretzky's theorem on nearly Euclidean subspaces

The Dvoretzky theorem (cf. Section X) is a purely finite-dimensional fact. It is essentially equivalent to the following geometric statement: Given k natural and $\varepsilon > 0$, there is an $n(k, \varepsilon) > 0$ such that if $n > n(k, \varepsilon)$ then every absolutely convex body in $\mathbf{R}^n$ (resp. $\mathbf{C}^n$) has a k-dimensional hyperplane section through the origin which is $(1+\varepsilon)$-spherical, i.e., the ratio of the radii of the Euclidean balls centred at the origin, circumscribed on the section and inscribed into it is $\leqslant 1+\varepsilon$. A satisfactory quantitative estimate of the function $(k, \varepsilon)\to n(k, \varepsilon)$ is due to Milman (1971):

(DV) *There is a function* $\varepsilon\to c(\varepsilon)$ *for* $\varepsilon > 0$ *such that every n-space has a subspace E with* $\dim E = k \geqslant c(\varepsilon)\log n$ *and* $d(E, l_k^2) < 1+\varepsilon$.

The estimate is sharp (for fixed ε):

If E is a subspace of l_n^∞ with $d(E, l_k^2) \leqslant 2$ then $k \leqslant c\log n$ where c is an absolute constant.

The above observation contrasts with the following

THEOREM (Figiel, Lindenstrauss and Milman, 1977). *There is an absolute constant $c > 0$ such that for every $\tau \geqslant 1$ every n-space whose cotype q constant is $\leqslant \tau$ has a subspace E with $k = \dim E \geqslant c\tau^{-2} n^{2/q}$ and $d(E, l_k^2) \leqslant 2$.*

Call a subspace $E \subset X$ *c-large* $(0 < c < 1)$ if $\dim E \geqslant c \dim X$. The theorem in the particular case $q = 2$ says that every n-space with cotype 2 constant $\leqslant \tau$ has $c\tau^{-2}$-large 2-Euclidean subspaces. Independently Kashin proved an even more surprising fact:

THEOREM (Kashin, 1977). *There is an absolute constant $c > 1$ such that for all n there is a subspace E of l_n^1 with* $\dim E = [n/2]$ *such that $\|x\|_1 \leqslant \sqrt{n}\|x\|_2 \leqslant c\|x\|_1$ for every x in E and for every x in the orthogonal complement of E with respect to the usual scalar product in $\boldsymbol{R}^n$ (resp. $\boldsymbol{C}^n$).*

Szarek (1978) found an elegant proof of this result; it uses the estimation of the volume ratio of the unit balls of l_n^1 and l_n^2. Further generalizations were obtained by Szarek and Tomczak-Jaegermann (1980).

Recently Johnson and Schechtman (1983) have shown that if $1 \leqslant p < q \leqslant 2$ then for every $\varepsilon > 0$ there is an absolute constant $c_{p,q}^{\varepsilon}$ such that l_n^p has a $c_{p,q}^{\varepsilon}$-large subspace E_n with $d(E_n, l_{\dim E_n}^q) < 1+\varepsilon$. Pisier (1983) generalized this result, replacing l_n^p's by any sequence of n-spaces whose so-called p-stable types behave as p-stable types of l_n^p.

All these results on the existence of "large" subspaces are obtained by more or less probabilistic methods. No constructive examples are known!

For a comprehensive discussion of nearly l_k^p $(1 \leqslant p \leqslant \infty)$ subspaces of finite-dimensional spaces, as well as of various techniques used with regard to this subject we refer to surveys by Maurey (1983) and Figiel (1984). The paper by Figiel, Lindenstrauss and Milman (1977) plays a fundamental role.

Dvoretzky's theorem has a few applications to infinite-dimensional spaces. It was used to show that a Banach space whose all subspaces are complemented is isomorphic to a Hilbert space (Lindenstrauss and Tzafriri, 1971) and that for every $\varepsilon > 0$ every separable infinite-dimensional Banach space has a total fundamental biorthogonal system bounded by $(1+\varepsilon)$ (Pełczyński, 1976). It was Dvoretzky (1961) who already observed that his theorem yields the following characterization of Hilbert spaces: an infinite-dimensional Banach space X is isometric to a Hilbert space iff for some $k \geqslant 2$ all k-dimensional subspaces of X are isometric. The following finite-dimensional analogue of this result is not yet solved:

Let $n > k \geqslant 2$. Is an n-space isometric to a Hilbert space assuming that all k-subspaces of the space are isometric? Equivalently, is an absolutely convex body in $\boldsymbol{R}^n$ (resp. $\boldsymbol{C}^n$) *an ellipsoid assuming that all k-dimensional sections of the body through the origin are affinely equivalent?*

The problem goes back to Banach and Mazur, cf. Banach (1932), p. 214. The answer is known and is "yes" in the following cases: real spaces — k even, $k < n$; k odd, $k < n-1$, complex spaces — k even, $k < n$; k odd, $2k \leqslant n$. The result is due to Gromov (1966), who used the algebraic topology technique of principal fibre bundles. The real case $n = 3$, $k = 2$ goes back to Auerbach, Mazur and Ulam (1935). The simplest open case is $n = 4$, $k = 3$.

XVI. The Banach–Mazur distance between n-spaces

The Banach–Mazur distance is a multiplicative metric (i.e., $\log d(\cdot, \cdot)$ is a usual metric) on the space B_n of all classes of isometries on n-spaces. Under this metric B_n is a compact space called the n-th Banach–Mazur compact. In this section we present some results concerning the diameter and the centre of B_n.

We begin with the following

THEOREM (John, 1948). *If $\ni$ is the ellipsoid of maximal volume inscribed in an absolutely convex body W in $\boldsymbol{R}^n$* (resp. $\boldsymbol{C}^n$) *then $\ni \subset W \subset \sqrt{n}\, \ni$.*

COROLLARY 1. *For every $E \in B_n$, $d(E, l_n^2) \leqslant \sqrt{n}$.*

Note that $d(l_n^\infty, l_n^2) = \sqrt{n}$, hence the inequality is sharp. Obviously Corollary 1 yields

COROLLARY 2. *$d(E,F) \leqslant n$ for all $E \in B_n$ and $F \in B_n$.*

What is the exact value of the diameter of B_n? Stromquist (1981) proved that the diameter of real B_2 is $3/2$; the unique pair of real 2-spaces whose Banach–Mazur distance is $3/2$ consists of l_2^∞ and the 2-space whose unit ball is the regular hexagon. This substantiates a conjecture of Asplund (1960). The diameter of B_3 is unknown. The right question to ask seems to be: what is the asymptotic order of the diameter of B_n as $n \to \infty$? Solving the problem which was open for a long time, E. D. Gluskin ingeniously proved the following

THEOREM (Gluskin, 1981). *There exists a $c > 0$ such that for $n = 1, 2, \ldots$ there are n-spaces X_n and Y_n with*

$$d(X_n, Y_n) > cn. \tag{4}$$

Briefly about the proof. The spirit is probabilistic. A class A_n of "random" n-spaces is constructured; a vast majority of pairs of the spaces satisfy (4). A_n consists of all n-spaces $(\boldsymbol{R}^n, |\cdot|_A)$ where $A = A(x_1, x_2, \ldots, x_{4n})$ is an absolutely convex body generated by the coordinate unit vectors and $4n$ "random" points $x_k \in S_{n-1} = \{x \in \boldsymbol{R}^n: \|x\|_2 = 1\}$ $(k = 1, 2, \ldots, 4n)$ and $|\cdot|_A$ denotes the gauge functional of A. In the sequel we identify A with the space $(\boldsymbol{R}^n, |\cdot|_A)$. The probability $\boldsymbol{P}_n$ on A_n is the transport via the surjection $(x_1, x_2, \ldots, x_{4n}) \to A(x_1, x_2, \ldots, x_{4n})$ of the Haar measure of the product S_{n-1}^{4n}. Note that on B_n there is no natural probability; in this way the difficulty is overcome!

Let $\|T\|_{A,B} = \|T: A \to B\|$ denote the operator norm of a $T: \boldsymbol{R}^n \to \boldsymbol{R}^n$. For $a > 0$ put

$$\mathscr{U}(a, n) = \{(A, B) \in A_n \times A_n: \exists T: \boldsymbol{R}^n \to \boldsymbol{R}^n,\ \|T\|_{A,B} < a\sqrt{n},\ \det T = 1\}$$

($\det T$ is the determinant of the matrix representing T in the natural unit vector basis). Next for a $B \in A_n$ we put

$$\mathscr{U}_B(a, n) = \{A \in A_n: (A, B) \in \mathscr{U}(a, n)\}. \tag{$*$}$$

The essence of the proof is to show

$$\exists a > 0: \lim_{n=\infty} \boldsymbol{P}_n\big(\mathscr{U}_B(a, n)\big) = 0 \quad \text{uniformly for } B. \tag{$^*_*$}$$

Clearly ($^*_*$) implies that, for large n, $\boldsymbol{P}_n \times \boldsymbol{P}_n\big(\mathscr{U}(a, n)\big) < 0.5$. Hence there are $A, B \in A_n$ such that $(A, B) \notin \mathscr{U}(a, n)$ and $(B, A) \notin \mathscr{U}(a, n)$. Thus, for every $S: \boldsymbol{R}^n \to \boldsymbol{R}^n$ and $T \in \boldsymbol{R}^n \to \boldsymbol{R}^n$ with $\det S = \det T = 1$, $\|T\|_{A,B}\|S\|_{B,A} \geqslant a^2 n$; in particular this holds for $S = T^{-1}$. Hence $d(A, B) \geqslant a^2 n$.

The delicate proof of ($^*_*$) is based on an estimation of the cardinality of ε-nets of some sets of operators in the space $B(l_n^2)$. In the final stage it reduces to a precise estimation from above of the ratio of the volumes of the unit ball of $B(l_n^2)$ and the unit ball of $l^2_{n^2}$ (cf. Saint-Raymond [a], for the evaluation of the first volume).

The results of John and Gluskin indicate a possibility that l_n^2 is asymptotically the unique centre of B_n. Precisely:

Given a sequence X_n with $X_n \in B_n$ $(n = 1, 2, \ldots)$ and $\lim_n d(l_n^2, X_n) = +\infty$. Does there exist a sequence Y_n with $Y_n \in B_n$ $(n = 1, 2, \ldots)$ such that $\lim_n d(X_n, Y_n) n^{-1/2} = +\infty$?

The answer is unknown even for $X_n = l_n^p$ $(p \neq 2)$.

Recently Tomczak-Jaegermann (1984) proved that if X and Y are symmetric n-spaces then $d(X, Y) \leqslant 2^{20} n^{1/2}$. (An n-space is symmetric if it has a basis such that all permutations of the elements of the basis and their multiplication by ± 1 extend to isometries of the space.) The same paper contains a generalization to symmetric bases of the classical fact that $d(l_n^p, l_n^q) = O(n^{a(p,q)})$, where $a(p,q) = \min(\frac{1}{2}, |1/p - 1/q|)$ for $1 \leqslant p$, $q \leqslant \infty$ (Gurarii, Kadec and Macaev, 1966). For other results on the Banach–Mazur distance we refer the reader to Benyamini and Gordon (1981), Lewis (1978), Pełczyński and Schütt (1981), Tomczak-Jaegermann (1978).

Powerful Gluskin's technique has recently been used to obtain sharp estimates for so-called basis constants of n-spaces (Gluskin, 1981a; Szarek, 1983; cf. also the report of Pisier, 1984), and symmetry constants of n-spaces (Mankiewicz, 1984). The sharp estimates for factorization constants of n-spaces through spaces with unconditional bases (and therefore, by a result of Lindenstrauss (1972), through symmetric spaces) were earlier obtained by Figiel, Kwapień and Pełczyński (1977) (cf. also Figiel and Johnson, 1980).

XVII. Some omitted topics

One can view Banach spaces as a mountain range. We have done a few excursions there. There are others equally beautiful. We shall now mention some topics of the theory of Banach spaces in which an essential progress has been achieved in the last decade but which are not discussed in this survey. The references are far from complete.

1. Theory of vector measures, integral representations of convex sets, the Radon–Nikodym and the Krein–Milman properties. A book by Diestel and Uhl (1977) presents the state of art until 1977. For more recent information see Diestel and Uhl (1983) and the lecture notes of Bourgin (1983) with a comprehensive bibliography; let us mention also a very recent paper by Bourgain and Pisier [a], which is not in the bibliography.

2. Some aspects of the geometry of the unit ball of a Banach space. (i) *Uniform convexity and smoothness.* The book of Lindenstrauss and Tzafriri (1979) and the lecture notes of Diestel (1975) and Bourgin (1983) cover the basic material. We also signalize a recent paper by Davis, Garling and Tomczak-Jaegermann (1984) in which the concept of complex modulus of convexity (introduced by Globevnik, 1975) is studied.

(ii) *Renorming theorems.* Diestel's lecture notes mentioned above presents the state of affairs until 1975. Of the many recent papers we only mention Davis, Ghoussoub and Lindenstrauss (1981), Fabian, Whitfield and Zizler (1983), Godefroy, Troyanski, Whitfield and Zizler (1983) and [a].

(iii) *Other topics*: a survey by Edgar and Wheeler [a] with an extensive list of references, papers by Godefroy (1981), Kunen and Rosenthal (1982), Larman and Phelps (1979), Ruess and Stegall (1982).

We do not feel competent to provide a suitable selection of references concerning the theory of non-separable Banach spaces, in particular weakly compactly generated (=WCG) spaces and their relatives, non-separable $C(S)$ and $L'(\mu)$ spaces, and various examples of non-separable spaces with curious properties. There is a need for a synthetic treatment of this subject.

References

Aldous D. J., 1979, Unconditional Bases and Martingales in $L_p(F)$, *Math. Proc. Cambridge Philos. Soc.* **85**, pp. 117–123.

Aldous D. J. 1981, Subspaces of L^1 via Random Measures, *Trans. Amer. Math. Soc.* **267**, pp. 445–463.

Aleksandrov A. B., 1982, The Existence of Inner Functions on the Ball (in Russian), *Mat. Sbornik* **118** (2), pp. 147–163.

Alspach D., 1978, Quotients of $C[0, 1]$ with Separable Dual, *Israel J. Math.* **29**, pp. 381–384.

Araujo A. and Gine E., 1980, *The Central Limit Theorem for Real and Banach-Valued Random Variables*, J. Wiley and Sons, New York.

Asplund E., 1960, Comparison between Plane Symmetric Convex Bodies and Parallelograms, *Math. Scand.* **8**, pp. 171–180.

Auerbach H., Mazur S. and Ulam S., 1935, Sur une propriété caractéristique de l'ellipsoïde, *Monatshefte für Mathematik und Physik* **42**, pp. 45–48.

Banach S., 1932, *Théorie des operations linéaires*, Monografie Matematyczne, Warszawa.

Banach S. and Mazur S., 1933, Zur Theorie der linearen Dimension, *Studia Math.* **4**, pp. 100–112.

Beck A., 1962, Convexity Condition in Normed Linear Spaces and the Strong Law of Large Numbers, *Proc. Amer. Math. Soc.* **13**, pp. 329–334.

Bennett G., Schur Multipliers, *Duke Math. J.* **44**, pp. 603–639.

Benyamini Y., 1978, An Extension Theorem for Separable Banach Spaces, *Israel J. Math.* **29**, pp. 24–30.

Benyamini Y. and Gordon Y., 1981, Random Factorization of Operators between Banach Spaces, *Journal d'Analyse Mathématique* **39**, pp. 45–74.

Benyamini Y., Rudin M. E. and Wage M., 1977, Continuous Images of Weakly Compact Subsets of Banach Spaces, *Pacific J. Math.* **70**, pp. 309–324.

Bessaga C. and Pełczyński A., 1958, On Bases and Unconditional Convergence of Series in Banach Spaces, *Studia Math.* **17**, pp. 151–164.

Bessaga C. and Pełczyński A., 1960, Banach Spaces of Continuous Functions IV, *Studia Math.* **19**, pp. 53–62.

Bochkarev S. V., 1974, Existence of a Basis in the Space of Functions Analytic in the Disc and Some Properties of the Franklin System (in Russian), *Mat. Sb. N. S.* **95** (137), pp. 3–18.

Bourgain J., 1981, *New Classes of $\mathcal{L}^p$-Spaces*, Lecture Notes in Math. 889, Springer-Verlag, Berlin–Heidelberg–New York.

Bourgain J., 1983, Embedding L^1 in L^1/H^1, *Trans. Amer. Math. Soc.* **278**, pp. 689–702.

Bourgain J., 1983a, New Banach Space Properties of the Disc Algebra and H^∞, *Acta Math.*, to appear.

Bourgain J., 1983b, The Non-Isomorphism of H^1 Spaces in a Different Number of Variables, *Bull. Soc. Math. Belgique, Ser. B*, **35** (II), pp. 127–136.

Bourgain J., 1983c, Some Remark on Banach Spaces in Which Martingale Difference Sequences are Unconditional, *Arkiv för Matematik* **21**, pp. 163–168.

Bourgain J., 1984, New Banach Space Properties of Certain Spaces of Analytic Functions, these *Proceedings*, pp. 945–951.

Bourgain J., [a], Extension of H_∞-Valued Operators and Bounded Bianalytic Functions (Institut Mittag-Leffler, Report No 6, 1983), *Trans. Amer. Math. Soc.*, to appear.

Bourgain J., [b] *Extension of a Result of Benedek, Calderon and Panzone* (preprint, Vrije Universiteit Brussel), *Arkiv för Matematik*, to appear.

Bourgain J. and Delbaen F., 1980, A Class of Special $\mathcal{L}_\infty$-Spaces, *Acta Math.* **145**, pp. 155–176.

Bourgain J., Fremlin D. H. and Talagrand M., 1978, Pointwise Compact Sets of Baire Measurable Functions, *Amer. J. Math.* **100**, pp. 845–886.

Bourgain J. and Pisier G., [a], A Construction of $\mathcal{L}_\infty$-Spaces and Related Banach Spaces, preprint, to appear.

Bourgain J., Rosenthal H. P. and Schechtman G., 1981, An Ordinal L^p-Index for Banach Spaces, with Application to Complemented Subspaces of L^p, *Ann. of Math.* **114**, pp. 193–228.

Bourgain J., and Rosenthal H. P., 1983, Applications of Theory of Semi-Embeddings to Banach Space Theory, *J. Functional Anal.* **52**, pp. 149–188.

Bourgain J. and Talagrand M., 1981, Dans un espace de Banach réticule solide, la propriété de Randon–Nikodym et celle de Krein–Milman sont équivalentes, *Proc. Amer. Math. Soc.* **81**, pp. 93–96.

Bourgin D., 1983, *Geometric Aspects of Convex Sets with the Randon–Nikodym Property*, Lecture Notes in Math. 993, Springer-Verlag, Berlin, Heidelberg, New York.

Bretagnolle J., Dacunha-Castelle D. and Krivine J. L., 1966, Lois stable et espaces L^p, *Ann. Inst. H. Poincaré* **2**, pp. 231–259.

Bukhvalov, A. V., Veksler A. I. and Lozanovskii G. Ya., 1979, Banach Lattices — Some Banach Space Aspects of the Theory (in Russian), *Uspekhi Mat. Nauk* **34** (2) (206), pp. 137–138.

Burkholder D. L., Martingale Transforms, *Ann. Math. Statist.* **37**, pp. 1494–1504.

Burkholder D. L., 1981, A Geometrical Characterization of Banach Spaces in which Martingale Difference Sequences Are Unconditional, *Ann. of Probability* **9**, pp. 997–1011.

Burkholder D. L., 1982, A Geometric Condition that Implies the Existence of Certain Singular Integrals of Banach-Space Valued Functions. In: Beckner W., Calderon A. P., Fefferman R. and Jones P. W., (eds.), *Conference on Harmonic Analysis in Honor of Antoni Zygmund*, Vol. I, Wadsworth Inc., Belmont, Calif., pp. 270–286.

Carleson L., 1980, An Explicit Unconditional Basis in H^1, *Bull. des Sciences Math.* **104**, pp. 405–416.

Casazza P. G., [a], Finite-Dimensional Decompositions in Banach Spaces. In: *Proc.*

Conference on the Occasion of M. M. Day Retirement, Lecture Notes, Springer-Verlag, Berlin–Heidelberg–New York.

Casazza P. G., Johnson W. B. and Tzafriri L., 1984, On Tsirelson's Space, *Israel J. Math.* **47**, pp. 81–98.

Ciesielski Z., 1966, Properties of the Orthonormal Franklin System II, *Studia Math.* **27**, pp. 289–323.

Day M. M., 1973, *Normed Linear Spaces*, third edition, Springer-Verlag, Berlin–Heidelberg–New York.

Davis W. J., Garling D. J. H. and Tomczak-Jaegermann N., 1984, The Complex Convexity of Quasi Normed Linear Spaces, *J. Functional Anal.* **55**, pp. 110–150.

Davis W. J., Ghoussoub N. and Lindenstrauss J., 1981, A Lattice Renorming Theorem and Applications to Vector-Valued Processes, *Trans. Amer. Math. Soc.* **263** (2), pp. 531–540.

Diestel J., 1975, *Geometry of Banach Spaces — Selected Topics*, Lecture Notes in Math., Springer-Verlag, Berlin–Heidelberg–New York.

Diestel J., and Uhl J. J., 1977, Vector Measures, *Mathematical Surveys*, AMS, Providence, R.I.

Diestel J. and Uhl J. J., 1983, Progress in Vector Measures — 1977–83, In: *Measures Theory and Its Applications, Proc. Conf. Sherbrooke 1982*, Lecture Notes in Math., Springer-Verlag, Berlin–Heidelberg–New York, pp. 144–192.

Dor L., 1975, On Sequences Spanning a Complex l_1-Space, *Proc. Amer. Math. Soc.* **47**, pp. 515–516.

Dunford N. and Schwartz J., 1958, *Linear Operators*, Vol. I, Interscience, New York.

Dvoretzky A., 1959, A Theorem on Convex Bodies and Applications to Banach Spaces, *Proc. Nat. Acad. Sci. (USA)* **45**, pp. 223–226.

Dvoretzky A., 1961, Some Results on Convex Bodies and Banach Spaces, In: *Proc. Symp. Linear Spaces, Jerusalem*, pp. 123–160.

Edgar G. A. and Wheeler R. F., [a], *Topological Properties of Banach Spaces*, preprint, to appear.

Enflo P., 1973, A Counterexample to the Approximation Property, *Acta Math.* **130**, pp. 309–317.

Enflo P., Lindenstrauss J. and Pisier G., 1975, On the Three Space Problem, *Math. Scand.* **36**, pp. 199–210.

Etcheberry A., 1975, Isomorphism of Spaces of Bounded Continuous Functions, *Studia Math.* **53**, pp. 103–127.

Farahat J., 1974, *Espaces de Banach contenant l_1, d'apres H. P. Rosenthal*, Séminaire Maurey–Schwartz, École Polytechnique (1973/74).

Fabian M., Whitfield J. H. M. and Zizler V., 1983, Norms with Locally Lipschitzian Derivatives, *Israel J. Math.* **44**, pp. 262–276.

Figiel T., 1984, Local Theory of Banach Spaces and Some Operator Ideals, these *Proceedings*, pp. 961–976.

Figiel T. and Johnson W. B., 1974, A Uniformly Convex Banach Space which Contains No l_p, *Compositio Math.* **29**, pp. 179–190.

Figiel T. and Johnson W. B., 1980, Large Subspaces of l_∞^n and Estimates of the Gordon–Lewis Constant, *Israel J. Math.* **37**, pp. 92–112.

Figiel T., Johnson W. B. and Tzafriri L., 1975, On Banach Lattices and Spaces Having Local Unconditional Structure with Applications to Lorentz Function Spaces, *J. Approximation Theory* **13**, pp. 395–412.

Figiel T., Kwapień S. and Pełczyński A., 1977, Sharp Estimates for the Constants of Local Unconditional Structure of Minkowski Spaces, *Bull. Acad. Polon. Sci. Ser. Sci. math. astr. et phys.* **25**, pp. 1221–1226.

Figiel T., Lindenstrauss J. and Milman V. D., 1977, The Dimension of almost Spherical Sections of Convex Bodies, *Acta Math.* **139**, pp. 53–94.

Figiel T. and Tomczak-Jaegermann N., 1979, Projections onto Hilbertian Subspaces of Banach Spaces, *Israel J. Math.* **33**, pp. 155–171.

Ghoussoub N. and Rosenthal H. P., [a], *Martingales, G_δ-Embeddings and Quotients of l_1* (preprint the University of Texas of Austin), to appear.

Giesy D. P., 1966, On a Convexity Condition in Normed Linear Spaces, *Trans. Amer. Math. Soc.* **125**, pp. 114–146.

Globevnik J., 1975, On Complex Strict and Uniform Convexity, *Proc. Amer. Math. Soc.* **47**, pp. 175–178.

Gluskin E. D., 1981, Diameter of Minkowski Compactum Approximatively Equals to n (in Russian), *Funktsional. Analiz i Prilozh.* **15** (1), pp. 72–73.

Gluskin E. D., 1981a, Finite-Dimensional Analogs of Spaces without Basis (in Russian), *Dokl. Akad. Nauk SSSR* **261** (5), pp. 1046–1050.

Godefroy G., Points de Namioka espaces normants applications a la théorie isométrique de la dualité, *Israel J. Math.* **38**, pp. 209–220.

Godefroy G., Troyanski S., Whitfield J. H. M. and Zizler V., 1983, Smoothness in Weakly Compactly Generated Banach Spaces, *J. Functional Analysis* **52**, pp. 344–352.

Godefroy G., Troyansky S., Whitfield J. H. M., and Zizler V., [a], *Locally Uniformly Rotund Renorming and Injections into $c_0(\Gamma)$*, Sonderforschungsbereich, to appear.

Gordon Y. and Lewis D. R., 1974, Absolutely Summing Operators and Locally Unconditional Structures, *Acta. Math.* **133**, pp. 27–48.

Gromov M. L., 1967, On a Geometric Conjecture of Banach (in Russian), *Izv. Akad. Nauk SSSR, ser. Mat.* **31**, pp. 1105–1114.

Grothendieck A., 1956, Résumé de la théorie métrique des produits tensoriels topologiques, *Boll. Soc. Mat. Sao Paulo*, pp. 1–79.

Guerre S. and Levy M., 1983, Espaces l^p dans les sous-espaces de L^1, *Trans. Amer. Math. Soc.*

Gurarii V. I., Kadec M. I. and Macaev V. I., 1966, Distance between Finite-Dimensional Analogs of l_p-Spaces (in Russian), *Mat. Sb.* **70** (112), pp. 24–29.

Haagerup U., 1981, *The Grothendieck Inequality for Bilinear Forms on C^*-Algebras*, Preprint, Matem. Inst. Odense Universitet No 1, to appear in *Advanced Math.*

Haagerup U., 1983, Solution of the Similarity Problem for Cyclic Representations of C^*-Algebra, *Ann. of Math.* **118**, pp. 215–240.

Hoffmann-Jørgensen J., 1972, *Sums of Independent Banach Space Valued Random Variables*, Aarhus Universitet preprint No 15.

Hoffmann-Jørgensen J., 1974, Sums of Independent Banach Space Valued Random Variables, *Studia Math.* **52**, pp. 159–186.

Hoffmann-Jørgensen J., 1977, Probability in Banach Spaces, École d'Ete' de Probabilites de Saint-Flour VI–1976, Lecture Notes in Math. No 598, Springer-Verlag, Berlin–Heidelberg–New York, pp. 1–186.

Hoffman-Jørgensen J. and Pisier G., 1976, The Law of Large Numbers and the Central Limit Theorem in Banach Spaces, *Ann. Prob.* **4**, pp. 587–599.

James R. C., 1950, Bases and Reflexivity of Banach Spaces, *Ann. of Math.* **52**, pp. 518–527.

James R. C., 1951, A Non-Reflexive Banach Space Isometric with Its Second Dual, *Proc. Nat. Acad. Sci. (USA)* **37**, pp. 174–177.

James R. C., 1964, Uniformly Non-Square Banach Spaces, *Ann. of Math.* **80**, pp. 542–550.

James R. C., 1978, Non-Reflexive Spaces of Type 2, *Israel J. Math.* **30**, pp. 1–13.

John F., 1948, *Extremum Problems with Inequalities as Subsidiary Conditions*, R. Courant Anniversary Volume, Interscience, New York, pp. 87–204.

Johnson W. B., 1980, Banach Spaces all of Whose Subspaces Have the Approximation Property, *Seminaire d'Analyse Fonctionalle 1979/1980*, exposé no 16, École Polytechnique, Paris.

Johnson W. B., Lindenstrauss J. and Schechtman G., 1980, On the Relation between Several Notions of Local Unconditional Structure, *Israel J. Math.* **37**, pp. 120–129.

Johnson W. B., König H., Maurey B. and Retherford J. R., 1979, Eigenvalues of p-Summing and l_p-Type Operators in Banach Spaces, *J. Functional Analysis* **32**, pp. 353–380.

Johnson W. B., Maurey, B., Schechtman G. and Tzafriri L., 1979, Symmetric Structures in Banach Spaces, *Memoirs Amer. Math. Soc.* **217**.

Johnson W. B. and Rosenthal H. P., 1972, On w^*-Basic Sequences and Their Applications to the Study of Banach Spaces, *Studia Math.* **43**, pp. 77–92.

Johnson W. B., Rosenthal H. P., and Zippin M., 1971, On Bases, Finite Dimensional Decompositions and Weaker Structures in Banach Spaces, *Israel J. Math.* **9**, pp. 488–506.

Johnson W. B. and Schechtman G., 1983, Embedding l_p^m into l_1^n, *Acta Math.* **149**, pp. 71–85.

Kadec M. I. and Pełczyński A., 1962, Bases Lacunary Sequences and Complemented Subspaces in the Spaces L_p, *Studia Math.* **21**, pp. 161–176.

Kaiser R. and Retherford J. R., 1983, Eigenvalue Distribution of Nuclear Operators: A Survey. In: W. Ruess (ed.), *Proc. Conf. Vector Measures and Functional Analysis*, Vorlesungen Fachbereich Math., Univ. Essen, pp. 245–287.

Kalton N. and Peck T., 1979, Twisted Sums of Sequence Spaces and the Three Spaces Problem, *Trans. Amer. Math. Soc.* **255**, pp. 1–30.

Kashin B. S., 1977, Diameter of Some Finite Dimensional Sets and Some Classes of Smooth Functions (in Russian), *Izv. Akad. Nauk. SSSR. Ser. Mat.* **41**, pp. 334–351.

Khmieleva T. E., 1981, Isomorphisms of Spaces of Bounded Continuous Functions (in Russian), *Zap. Nauchn. Sem. Leningrad, Otdel. Mat. Inst. Steklov (LOMI)* **113**, pp. 243–246.

Kislyakov S. V., 1976, On Spaces with "Small" Annihilators (in Russian), *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI)* **65**, pp. 192–195.

Kislyakov S. V., 1977, *The Space of Continuously Differentiable Functions on the Torus Has no Local Unconditional Structure* (in Russian), Preprint LOMI, P-I-77, Leningrad.

König H., 1977, Operator Properties of Sobolev Embeddings over Unbounded Domains, *J. Functional Analysis* **24**, pp. 32–51.

König H., Retherford J. R., and Tomczak-Jeagermann N., 1980, On the Eigenvalues

of (p. 2)-Summing Operators and Constants Associated to Normed Spaces, *J. Functional Analysis* **37**, pp. 88–126.
Krivine J. L., 1974, *Théorèmes de factorisation dans les espaces réticules*, Séminaire Maurey-Schwartz 1973/74, Exposés 22–23, École Polytechnique, Paris.
Krivine J. L., 1976, Sous-espaces de dimension finite des espaces de Banach réticules, *Ann. of Math.* **104**, pp. 1–29.
Krivine J. L., 1979, Constantes de Grothendieck et fonctions de type positif sur les sphères, *Advances of Math.* **31**, pp. 16–30.
Krivine J. L. and Maurey B., 1981, Espaces de Banach stables, *Israel J. Math.* **10**, pp. 379–390.
Kunen K. and Rosenthal H. P., 1982, Martingale Proofs of Some Geometrical Results in Banach Space Theory, *Pacific J. Math.* **100**, pp. 153–175.
Kwapień S., 1970, On a Theorem of L. Schwartz and Its Applications to Absolutely Summing Operators, *Studia Math.* **38**, pp. 193–201.
Kwapień S., 1972, Isomorphic Characterizations of Inner Product Spaces by Orthogonal Series with Vector Valued Coefficients, *Studia Math.* **44**, pp. 583–595.
Kwapień S., 1972a, On Operators Factorizable through L^p-Spaces, *Bull. Soc. Math. France Memoire* **31/32**, pp. 215–229.
Kwapień S. and Pełczyński A., 1980, Absolutely Summing Operators and Translation Invariant Spaces of Functions on Compact Abelian Groups, *Math. Nachr.* **94**, pp. 303–340.
Lacey H. E., 1974, *The Isometric Theory of Classical Banach Spaces*, Springer-Verlag, Berlin–Heidelberg–New York.
Larman D. G. and Phelps R. R., 1979, Gateaux Differentiability of Convex Functions on Banach Spaces, *J. London Math. Soc.* **20**, pp. 115–127.
Levy P., 1937, *Théorie de l'addition de variables aléatoires*, Paris.
Lewis D. R., 1979, An Isomorphic Characterization of the Schmidt Class, *Compositio Math.* **30**, pp. 291–297.
Lewis D. R. and Stegall Ch., 1973, Banach Spaces whose Duals Are Isomorphic to $l_1(\Gamma)$, *J. Functional Analysis* **12**, pp. 177–187.
Lindenstrauss J., 1967, On Complemented Subspaces of m, *Israel J. Math.* **5**, pp. 153–156.
Lindenstrauss J., 1972, A Remark on Symmetric Bases, *Israel J. Math.* **13**, pp. 313–320.
Lindenstrauss J. and Pełczyński A., 1968, Absolutely Summing Operators in L_p Spaces and Their Applications, *Studia Math.* **29**, pp. 275–326.
Lindenstrauss J. and Tzafriri L., 1971, On the Complemented Subspaces Problem, *Israel J. Math.* **9**, pp. 263–269.
Lindenstrauss J., 1973, *Classical Banach Spaces*, Lecture Notes in Math., Springer-Verlag, Berlin–Heidelberg–New York.
Lindenstrauss J., 1977, Classical Banach Spaces I, Springer-Verlag, Berlin–Heidelberg–New York.
Lindenstrauss J., 1979, Classical Banach Spaces II, Springer-Verlag, Berlin–Heidelberg–New York.
E. Løw, 1982, A Construction of Inner Functions on the Unit Ball in C^p, *Invent. Math.* **67**, pp. 223–229.
Mankiewicz P., 1984, Finite Dimensional Banach Spaces with Symmetry Constant of Order $\sqrt{n}$, *Studia Math.* **79**, pp. 193–200.

Marcus M. and Pisier G., 1981, Random Fourier Series with Applications to Harmonic Analysis, *Annals of Math. Studies* **101**, Princeton, N. J.

Maurey B., 1974, Théoremès de factorisation pour les opérateurs linéaires à valeurs dans un espace L^p, *Asterisque* **11**, Soc. Math. France.

Maurey B., 1975, Quelques problèmes de factorisation d'opérateurs linéaires, *Proceedings ICM Vancouver 1974*, Vol. II, pp. 75–79.

Maurey B., 1975a, Systéme de Haar, *Séminaire Maurey-Schwartz 1974/75*, Exposés No 1–2, École Polytechnique, Paris.

Maurey B., 1983, Sous-espaces l^p des espaces de Banach, *Séminaire Bourbaki* **606**, Paris.

Maurey B. and Pisier G., 1976, Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach, *Studia Math.* **58**, pp. 45–90.

Maurey B. and Rosenthal H. P., 1977, Normalized Weakly Null Sequences with no Unconditional Subsequence, *Studia Math.* **61**, pp. 77–98.

Milman V. D., 1971, A New Proof of the Theorem of A. Dvoretzky on Sections of Convex (in Russian), *Funktsional. Analiz i Prilozhen.* **5**, pp. 28–37.

Milutin A. A., 1966, Isomorphisms of Spaces of Continuous Functions on Compacta of Power Continuum (in Russian), *Teor. Funktsii, Funktsional. Analiz i Prilozhen., Kharkov* **2**, pp. 150–156.

Mitjagin B. S., 1970, The Homotopy Structure of the Linear Group of a Banach Space (in Russian), *Uspekhi Mat. Nauk* **25** (5), pp. 63–106.

Mitjagin B. S. and Pełczyński A., 1967, Nuclear Operators and Approximative Dimension, *Proc. ICM, Moscow, 1966*, Moscow, pp. 366–372.

T. Odell, 1980, Applications of Ramsey Theorems to Banach Space Theory. In: H. Elton Lacey (ed.), *Notes in Banach spaces*, Austin, pp. 379–404.

Odell T. and Rosenthal H. P., 1975, A Double Dual Characterization of Separable Banach Spaces Containing l_1, *Israel J. Math.* **20**, pp. 375–384.

Orlicz W., 1933, Über unbedingte Konvergenz in Funktionenräumen I, II, *Studia Math.* **4**, pp. 33–37; 41–47.

Ørno P., 1976, A Note on Unconditionally Converging Series in L_p, *Proc. Amer. Math. Soc.* **59**, pp. 252–253.

Pełczyński A., 1967, A Characterization of Hilbert-Schmidt operators, *Studia Math.* **28**, pp. 355–360.

Pełczyński A., 1971, Any Separable Banach Space with the Bounded Approximation Property is a Complemented Subspace of a Banach Space with a Basis, *Studia Math.* **40**, pp. 239–242.

Pełczyński A., 1974, Sur certaines propriétés isomorphiques nouvelles des espaces de Banach de fonctions holomorphes A et H^∞, *C. R. Acad. Sci. Paris Sér. A* **279**, pp. 9–12.

Pełczyński A., 1976, All Separable Banach Spaces Admit for Every $\varepsilon > 0$ Fundamental Total Biorthogonal System Bounded by $1+\varepsilon$, *Studia Math.* **55**, pp. 295–304.

Pełczyński A., 1977, *Banach Spaces of Analytic Functions and Absolutely Summing Operators*, Regional conference series in mathematics, Vol. 30, Amer. Math. Soc.

Pełczyński A., 1980, Geometry of Finite Dimensional Banach Spaces and Operators Ideals. In: H. Elton Lacey (ed.), *Notes in Banach spaces*, Austin, pp. 81–181.

Pełczyński A. and Bessaga C., 1979, Some Aspects of the Present Theory of Banach Spaces. In: Stefan Banach, *Oeuvres*, Volume II, Warszawa, pp. 220–302.

Pełczyński A. and Rosenthal H. P., 1975, Localization Techniques in L^p-Spaces, *Studia Math.* **52**, pp. 263–289.

Pełczyński A. and Schütt C., 1981, Factoring the Natural Injection $l^{(n)}: L_n^\infty \to L_n^1$ through Finite Dimensional Banach Spaces and Goemetry of Finite Dimensional Unitary Ideals, *Advances in Math.*, *Supplementary Studies* **79**, pp. 653–682.

Pełczyński A. and Semadeni Z., 1959, Spaces of Continuous Functions III, *Studia Math.* **18**, pp. 211–222.

Peller V. V., 1983, Estimates of Functions of Power Bounded Operators on Hilbert Spaces, *Journal of Operator Theory* **7**, pp. 341–372.

Pietsch A., 1963, Zur Fredholmschen Theorie in lokalkonvexen Räumen, *Studia Math.* **22**, pp. 162–179.

Pietsch A., 1967, Absolut p-summierende Abbildungen in normierten Räumen, *Studia Math.* **28**, pp. 333–353.

Pietsch A., 1978, *Operator Ideals*, Berlin.

Pietsch A., 1980, Weyl Numbers and Eigenvalues of Operators in Banach Spaces, *Math. Annalen* **247**, pp. 149–178.

Pietsch A., *Operator Ideals* (in Russian), Izdatelstvo Mir, Moskva.

Pisier G., 1975, Un exemple concernant la super-réflexivité, *Séminaire Maurey–Schwartz 1974/75*, École Polytechnique, Paris.

Pisier G., 1978, Some Results on Banach Spaces without Local Unconditional Structure, *Compositio Math.* **37**, pp. 3–19.

Pisier G., 1978a, Une nouvelle classe d'espaces de Banach verifiant le Théorème de Grothendieck, *Annal. Inst. Fourier* **28**, pp. 69–90.

Pisier G., 1978b, Grothendieck's Theorem for Non-Commutative C^*-Algebras with an Appendix on Grothendieck's Constant, *J. Functional Analysis* **29**, pp. 397–415.

Pisier G., 1978c, *Séminaire sur la geometrie des espaces de Banach 1977/78*, Exposes 14, 17–18, École Polytechnique, Palaiseau.

Pisier G., 1979, Some Applications of the Complex Interpolation Method to Banach Lattices, *Journal d'Analyse Mathématique* **35**, pp. 265–281.

Pisier G., 1982, *Les produits tensoriels d'espaces de Banach* depuis Grothendieck Séminaire Initiation a l'Analyse G. Choquet M. Rogalski, J. Saint-Raymond, 21 année 1981/1982, Université Paris VI.

Pisier G., 1982a, Holomorphic Semi-Groups and the Geometry of Banach Spaces, *Ann. of Math.* **115**, pp. 379–392.

Pisier G., 1983, On the Dimension of the l_p^n-Subspaces of Banach Spaces for $1 \leq p < 2$, *Trans. Amer. Math. Soc.*

Pisier G., 1983a, Counterexamples to a Conjecture of Grothendieck, *Acta Math.* **151**, pp. 181–208.

Pisier G., 1984, Finite Rank Projections on Banach Spaces and a Conjecture of Grothendieck, these *Proceedings*, pp. 1027–1039.

Ringrose J. R., 1975, Operator Algebras and Abelian Subalgebras, *Proc. ICM Vancouver 1974*, pp. 105–110.

Rosenthal H. P., 1972, On Factors of $C[0, 1]$ with Non-Separable Dual, *Israel J. Math.* **13**, pp. 361–378.

Rosenthal H. P., 1973, On Subspaces of L^p, *Ann. of Math.* **97**, pp. 344–373.

Rosenthal H. P., 1974, A Characterization of Banach Spaces Containing l^1, *Proc. Nat. Acad. Sci (USA)* **71**, pp. 2411–2413.

Rosenthal H. P., 1976, Some Applications of p-Summing Operators to Banach Space Theory, *Studia Math.* **58**, pp. 21–43.
Rosenthal H. P., 1977, Pointwise Compact Subsets of the First Baire Class, *Amer. J. Math.* **99**, pp. 362–378.
Rosenthal H. P., 1978, On a Theorem of J. L. Krivine Concerning Block Finite-Representability of l^p in General Banach Spaces, *J. Funct. Anal.* **28**, pp. 197–225.
Rosenthal H. P., 1983, Sign-Embeddings of L^1. In: A.C. Plei and S.J. Sidney (eds.), *Banach Spaces, Harmonic Analysis and Probability Theory, Proceedings, 1980-1981*, Springer Lecture Notes 995, pp. 155–165.
Rosenthal H. P., [a], Embeddings of L^1 in L^1 (preprint), to appear.
Ruess W. H. and Stegall Ch. P., 1982, Extreme Points in Duals of Operator Spaces, *Math. Annalen* **261**, pp. 535–546.
Saint-Raymond J., [a], *Le volume des idéaux d'operateurs classiques*, Paris IV, preprint, *Studia Math.* **80**, to appear.
Schaefer H. H., 1974, Banach Lattices and Positive Operators, Springer-Verlag, Berlin–Heidelberg–New York.
Schechtman G., 1979, A Tree-Like Tsirelson Space, *Pacific J. Math.* **83**, pp. 523–530.
Schwartz L., 1969, Probabilités cylindriques et applications radonifiantes, *C.R. Acad. Paris* **268**, pp. 646–648.
Sém., 1972/73, *Séminaire Maurey–Schwartz*, École Polytechnique, Paris.
Sém., 1973/74, *Séminaire Maurey–Schwartz*, École Polytechnique, Paris.
Sém., 1974/75, *Séminaire Maurey–Schwartz*, École Polytechnique, Paris.
Sém., 1975/76, *Séminaire Maurey–Schwartz*, École Polytechnique, Paris.
Sém., 1977/78, *La géométrie de espaces de Banach*, École Polytechnique, Palaiseau.
Sém., 1978/79, *D'analyse fonctionnelle*, École Polytechnique, Palaiseau.
Sém., 1979/80, *D'analyse fonctionnelle*, École Polytechnique, Palaiseau.
Sém., 1980/81, *D'analyse fonctionnelle*, École Polytechnique, Palaiseau.
Sobczyk A., 1941, Projection of the Space m on Its Subspace c_0, *Bull. Amer. Math. Soc.* **47**, pp. 938–947.
Stromquist W., 1981, The Maximum Distance between Two-Dimensional Spaces, *Math. Scand.* **48**, pp .205–225.
Szankowski A., 1976, A Banach Lattice without the Approximation Property, *Israel J. Math.* **24**, pp. 329–337.
Szankowski A., 1978, Subspaces without Approximation Property, *Israel J. Math.* **30**, pp. 123–129.
Szankowski A., 1981, $B(H)$ Does not Have the Approximation Property, *Acta Math.* **147**, pp. 89–108.
Szarek S. J., 1978, On Kashin's almost Euclidean Orthogonal Decomposition of l_1^n, *Bull. Acad. Polon. Sci. Ser. mat. astr. et phys.* **26**, pp. 691–694.
Szarek S. J., 1983, The Finite Dimensional Basis Problem with an Appendix on Nets of Grassman Manifold, *Acta Math.* **151**, pp. 153–179.
Szarek S. and Tomczak-Jaegermann N., 1980, On Nearly Euclidean Decomposition for Some Classes of Banach Spaces, *Compositio Math.* **40**, pp. 367–385.
Talagrand M., 1983, La structure des espaces de Banach réticules ayant la propriété de Radon–Nikodym, *Israel J. Math.* **44**, pp. 213–220.
Tomczak-Jaegermann N., 1974, The Moduli od Smoothness and Convexity and the Rademacher Averages of Trace Classes S_p $(1 \leqslant p < \infty)$, *Studia Math.* **50**, pp. 163–182.

Tomczak-Jaegermann N., 1978, The Banach–Mazur Distance between the Trace Classes c_p^n, *Proc. Amer. Math. Soc.* **72**, pp. 305–308.

Tomczak-Jaegermann N., 1984, The Banach–Mazur Distance between Symmetric Spaces (preprint Texas A. and M. University), to appear in *Israel J. Math.*

Tonge A., 1976, Banach Algebras and Absolutely Summing Operators, *Math. Proc. Cambridge Phil. Soc.* **80**, pp. 465–473.

Tsirelson B. S., 1974, Not Every Banach Space Contains l_p or c_0 (in Russian), *Functional Anal. Appl.* **8** (2), pp. 57–60.

Varopoulos N. Th., 1974, On an Inequality of Von Neumann and an Application of the Theory of Tensor Products to Operator Theory, *J. Functional Analysis* **16**, pp. 83–100.

Wojtaszczyk P., 1982, The Franklin System is an Unconditional Basis in H^1, *Arkiv för Matematik* **20**, pp. 293–300.

Wojtaszczyk P., 1983, Hardy Spaces on the Complex Ball Are Isomorphic to Hardy Spaces on the Disc, $1 \leqslant p < \infty$, *Ann. of Math.* **118**, pp. 21–34.

Woyczyński W., 1978, Geometry and Martingales in Banach Spaces, Part II. In: J. Kuelbs (ed.), *Independent Increments, Probability on Banach Spaces*, Advances in Probability and Related Topics, Vol. 4, New York and Basel.

Zippin M., 1977, The Separable Extension Problem, *Israel J. Math.* **26**, pp. 372–387.

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
ŚNIADECKICH 8, 00-950 WARSZAWA, POLAND

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

DAVID RUELLE

Turbulent Dynamical Systems

Recent developments in the dynamical theory of the onset of turbulence are reviewed, and some historical and philosophical aspects of the subject are discussed.

Introduction

The phenomenon of hydrodynamic turbulence constitutes one of the great puzzles of theoretical physics, and deep studies have been devoted to it, over a long period of time. In recent years a conceptual clarification of some essential points has been obtained. Yet, other essential points remain obscure, and we do not have to-day a general theory of turbulence.

From an operational point of view it is easy enough to agree that certain flows are turbulent. Since the phenomenon is ubiquitous, and accessible to everyone for inspection, a number of theoretical interpretations have been proposed. Von Neumann [82] has given a lucid review of the early theories of turbulence. Many good ideas are contained in these theories, but they fail to explain the phenomenon. In brief, this failure can be understood as follows. The basic equations of hydrodynamics are nonlinear partial differential equations; their mathematical study is hard, and so is their numerical treatment in the turbulent regime. Experimental studies are also much more difficult than one would think, when precise quantitative information is desired.

The recent improvement of our understanding of the nature of turbulence has three different roots. The first is the injection of new mathematical ideas from the theory of dynamical systems. The second is the availability of powerful computers which permit, among other things, experimental mathematics on dynamical systems and numerical simulation of hydrodynamic equations. The third is the improvement of experimental techniques (in particular, Doppler measurements of velocities by use of a laser beam, and then numerical Fourier analysis of the time series obtained).

The discussion of hydrodynamic time evolution from the point of view of the qualitative theory of dynamical systems has had a profound impact. This was at first not due to the proof of any deep theorem, but rather to the reevaluation of accepted ideas. In the present review we shall analyze the somewhat complex evolution of ideas on the onset of turbulence, and the related problem of sensitive dependence on initial condition. Our discussion will thus emphasize historical and philosophical aspects.[1] I am aware of the difficulties of such an enterprise, but feel that the articulation of mathematics, physics and computer work in to-day's science deserves study, and that an imperfect discussion is better than no discussion at all.

Hydrodynamic time evolution as a dynamical system

A reasonable description of fluid dynamics is given by the *Navier–Stokes equation* (1) and the *incompressibility condition* (2) in d dimensions ($d = 3$ normally, but $d = 2$ is also much studied):

$$\frac{dv_i}{dt} = -\sum_{j=1}^{d} v_j \cdot \frac{\partial v_i}{\partial x_j} + \nu \Delta v_i - \frac{\partial p}{\partial x_i} + g_i, \tag{1}$$

$$\sum_j \frac{\partial v_j}{\partial x_j} = 0. \tag{2}$$

In these equations (v_i) is the velocity field of the fluid enclosed in a region Ω, ν is the viscosity coefficient, p the pressure (divided by the density), and (g_i) describes an external force. By projecting (1) on a space of divergence free vector fields, one takes (2) into account, and the pressure term is eliminated. The fluid sticks to the boundary. This boundary condition is imposed by writing $v_i = v_i^o = v_i^o + w_i$ where (w_i) belongs to a functional space $\mathscr{H}$ of vector fields vanishing on $\partial\Omega$. One has thus an evolution equation of the form

$$\frac{dw}{dt} = F_\mu(w), \tag{3}$$

where the right-hand side is assumed to be time independent, and μ is a parameter describing the intensity of external action exerted on the fluid through the force (g_i) or the boundary condition (v_i^o).

---

[1] For reviews of the same subject from a different viewpoint see Ruelle [69], [71].

The derivation of (1) involves a "linear response" approximation, and the Navier–Stokes equation is thus essentially less exact than the equations of celestial mechanics for instance. In the two-dimensional case ($d = 2$) there is a good existence and uniqueness theorem (Leray, Ladyzhenskaya). For $d = 3$, existence and uniqueness can be proved for small times, and a "weak solution", introduced by Leray[2] may have singularities and not be unique. In fact it is not known at present if singularities actually occur.[3] One knows, however, that the set of singularities cannot be too large. The study of this point has been initiated by Leray and pursued by Scheffer [76] and Caffarelli, Kohn and Nirenberg [5]. In [5] it is shown that the set of singularities in 4 dimensional space time has Hausdorff 1-dimensional measure 0. The low dimension of the singularities (they cannot form a curve) implies that they may not be very conspicuous, if they are present. It also implies that they probably do not have much to do with the physical phenomenon of turbulence (see below the discussion of intermittency).

The physical problem of turbulence, thus, is related in ways which have not yet become clear to the mathematical problem of understanding the Navier–Stokes equation. In what follows we shall take the region Ω to be bounded, and we shall *assume* that the evolution equation (3) defines a dynamical system. By this we mean that there is a bounded open set U in the space of square integrable w's, such that (3) has a good solution $f^t w$ for $t > 0$ and initial conditions $w \in U$, and $f^t w \in U$ for sufficiently large t. In particular we have the semigroup property

$$f^s \circ f^t = f^{s+t} \tag{4}$$

If $d = 2$ the existence of U can be proved (under suitable regularity conditions for $\partial\Omega$, (g_i) and (v_i^0)). When $d = 3$ it is not known if U exists in general. (It may then be physically required to replace the Navier–Stokes equation by another evolution equation, but we do not discuss this possibility here). If the dynamical system (f^t) exists, it has nice properties: $(t, w) \to f^t w$ is real analytic, and f^t is compact (it sends bounded sets to relatively compact sets, the derivative $D_w f^t$ is a compact linear map).[4]

---

[2] Leray [39] treated the case $\Omega = \boldsymbol{R}^3$, Hopf [24] later discussed the case of bounded Ω.

[3] The solution of a time evolution equation with "good" initial data may become "bad" in the sense that the evolution equation is no longer a useful physical approximation. Whether the problem occurs here or not is unknown but, as stressed by Leray, the question is of general relevance for the equations of mathematical physics.

[4] For Navier–Stokes theory see the monographs by Ladyzhenskaya [32], Lions [41], Serrin [77], Temam [81]. See also the excellent review by Foias and Temam [16], and Ruelle [70] where the dynamical systems viewpoint is discussed.

QUESTION. Can one define a dynamical system with weak solutions of the Navier–Stokes equation? In other words can (4) be made to survive when singularities are present ($d = 3$)?

Strange attractors

In equation (3) the parameter $\mu \geqslant 0$ is a measure of the external forces acting on the fluid. In practice μ is often the *Reynolds number* (or the *Rayleigh number* in convection experiments, where the temperature is introduced). If $\mu = 0$ (no external force) the fluid goes to rest, and for small μ tends to a steady state. For somewhat larger action, a periodic oscillation may appear. In terms of dynamical systems, the occurrence of periodic solutions is explained by the *Hopf bifurcation* [22]. Both Landau [33], [34] and Hopf [23] have suggested that, as μ is increased, more and more frequencies appear, and that a quasiperiodic motion is obtained:

$$w(t) = F(\omega_1 t, \dots \omega_k t)$$

where $\omega_1, \dots, \omega_k$ are the frequencies (for a certain value of μ) and F is periodic of period 2π in each argument separately. The quasiperiodic motion must occur on an attracting k-dimensional torus in the infinite dimensional phase space of the fluid (the open set U of the last section). When k is large enough, the time dependence of w is quite complicated, and Hopf and Landau propose that the fluid is turbulent. If this view is accepted, every independent frequency of the system requires one dimension of phase space. Furthermore the system does not depend sensitively on initial condition. This means that a small change δw_0 in initial condition does not grow exponentially with time.

From a mathematical point of view, there is something wrong with quasiperiodic motions on a k-torus, $k \geqslant 2$: these motions are non generic. This means that quasiperiodicity may be replaced by something else under a small perturbation of the evolution equation (3). Takens and myself [75] proposed in 1971 that "something else", which we called *strange attractors*, should describe turbulence. The prototype of strange attractors which we had in mind were the "Strange" Axiom A attractors introduced by Smale [80], and which can appear by perturbation of quasiperiodic motions on the k-torus for $k \geqslant 3$ (see [52], in [75] we needed $k \geqslant 4$). The advantage which we saw to strange attractors over quasiperiodicity is that a continuum of frequencies can be produced with a motion in finite dimensional space (thus involving a finite number of degrees of freedom only). In physical language, the nonlinear interaction between three "modes" already produces "continuous spectrum".

Although the ideas expressed in [75] were less completely new than we thought (see below), they were the first proposal, made explicitly and in print, of a general interpretation of hydrodynamic turbulence, in the framework of dynamical systems, rejecting the quasiperiodic dogma. The reaction of the scientific public to our proposal was quite cold. In particular, the notion that continuous spectrum would be associated with a few degrees of freedom was viewed as heretical by many physicists.[5] Finally, around 1974–1975, the problem was settled by the hydrodynamical experiments of Ahlers [1], and Gollub and Swinney [19], and the computer experiments of McLaughlin and Martin [48], [49]. These experiments (and many others which followed) show that continuous spectrum appears fairly suddenly hen the parameter μ (Rayleigh or Reynolds number) is increased. This is in agreement with the strange attractor picture, and contradicts the idea that new frequencies are added one after the other, as in Landau theory.

We must now come back to an important paper [42] published in 1963 by Lorenz in a meteorology journal, and which largely escaped the attention of mathematicians and physicists for a while. This paper became justly popular after a note by Guckenheimer [20] (published in 1976) brought attention to it. Lorenz does not discuss turbulence in general, but considers a simple evolution equation of the type (3), with $x \in \boldsymbol{R}^3$, which is a rough model for hydrodynamical convection. (As a meteorologist, Lorenz is interested in turbulent convection in the atmosphere). The numerical study of the Lorenz equation reveals a new strange attractor (not of Axiom A type). Sensitivity to initial condition is exhibited, and Lorenz takes argument of this to explain why meteorologists cannot accurately predict the weather long in advance.

It seems that some Russian mathematicians were also unhappy with Landau's quasiperiodic theory of turbulence. V. Arnold informs me that the subject was discussed in Moscow[6] and that he mentioned it in a seminar in Paris in 1965. Such preoccupations motivated Arnold's well-known paper [2]. Unfortunately, no non trivial result about the qualitative dynamics of Navier–Stokes was proved, and the ideas mentioned by Arnold remained unpublished. In fact, while the impact of the new mathematical

[5] As Monin [50] remarks in 1978: "it was tacitly assumed only ten years ago that only stationary points and closed or quasiperiodic orbits could be attractors for the phase paths".

[6] Letter dated 1980.

ideas on turbulence has completely changed the subject, we still have no hard theorem on the existence of strange attractors for the Navier–Stokes equation.

In the discussion (below) of sensitive dependence on initial condition we shall find other precursors of the modern ideas on turbulence.

The onset of turbulence

After the experiments of Ahlers, Gollub and Swinney, and the numerical work of McLaughlin and Martin mentioned earlier, a number of studies on the onset of turbulence followed. The onset of turbulence is the region of low values of μ (Reynolds or Rayleigh number) where the fluid starts to exhibit weak turbulence. It has now become apparent that a weakly turbulent viscous fluid behaves — as a dynamical system — like a generic dynamical system on a low dimensional manifold. Experimental studies (largely based on the frequency spectrum) exhibit periodicity, (with 2, sometimes 3 basic frequencies), strange attractors with sensitive dependence on initial condition, and some curious phenomena like the Feigenbaum bifurcation (see below).

One can prove (Mallet-Paret [43]) that, at finite μ, the Navier–Stokes time evolution is attracted asymptotically to a set of finite (Hausdorff) dimension. It is thus reasonable that weak turbulence appear finite dimensional. It may be more surprising that a viscous fluid behaves very much (from the dynamical viewpoint) like the solution of a randomly chosen equation (3) on a finite dimensional manifold.

Extensive computer studies of low dimensional dynamical systems have shown that sensitive dependence on initial condition is quite common, but mostly appears in systems for which we have no good mathematical theory (non Axiom A). We don't even have a very good mathematical definition of strange attractors. An *attractor* is a set such that the orbits of nearby points tend to it. This may be complemented by an irreducibility condition (see Ruelle [72]). The attractor is *strange* if it exhibits sensitive dependence on initial condition, i.e., exponential growth of small perturbations of initial condition. A precise definition involves the choice of an ergodic measure on the attractor (see below: ergodic theory). Unfortunately we do not know in general what ergodic measure to select.

One great success of the theory of the onset of turbulence is the observation of the *Feigenbaum bifurcation*. This is a new codimension 1-bifurcation first discovered numerically. As the bifurcation parameter μ is increased, an attracting periodic orbit of period T is successively replaced at values $\mu_1, \ldots, \mu_k, \ldots$ of μ by attracting orbits of period $\approx 2^k T$, and μ_k tends to μ_∞ so that $(\mu_k - \mu_\infty)/(\mu_{k+1} - \mu_\infty)$ tends to a universal constant. Beyond μ_∞, "chaotic" behavior with sensitive dependence on initial condition is observed (although not yet proved to exist in general). The Feigenbaum bifurcation is beautifully visible in the frequency analysis of the experimental data of Libchaber and Maurer [40] among others; it cannot possibly be mistaken for something else. The theory of the Feigenbaum bifurcation has started with the deep analysis of Feigenbaum [13], [14], [15] based on the physical idea of the renormalization group. This analysis involves looking for a fixed point in a functional space, and has been made rigorous by Landford's work [35] taken in conjunction with that of Collet, Eckmann and Koch [9] (see also Campanino, Epstein and Ruelle [7], [6]). Lanford's proof is remarkable in that it makes rigorous use of the computer to obtain numerical estimates which would be exceedingly painful to do by hand.

Many sequences of bifurcations lead to turbulence. Eckmann [12] calls them *scenarios*. Three main scenarios have been investigated. The quasiperiodic scenario (Ruelle, Takens and Newhouse [52], [75]), involves the creation of a quasiperiodic 2-torus and its destruction with appearance of a strange attractor. The period doubling cascade scenario is the Feigenbaum bifurcation. The intermittent[7] scenario of Pomeau and Manneville [60] corresponds to a saddle-node bifurcation and manifests itself as "turbulent bursts" in an apparently periodic background. The three scenarios have all been clearly recognized experimentally, but their theoretical study is still quite incomplete.

Strange attractors in general dissipative systems

Our discussion of the onset of hydrodynamic turbulence has made no use of the hydrodynamic equations. A "turbulent" behavior may therefore be expected in all kinds of natural systems. It is convenient to exclude here conservative (i.e. Hamiltonian) systems because of their special (non generic) character: conservative systems may show sensitive dependence on initial condition, but cannot have attractors because of the conservation of Liouville measure. We are thus left with the idea that dissipative (i.e. non conservative) systems exhibit turbulence. For instance, it is predicted that homogeneous chemical reactions may exhibit aperiodic

---

[7] The temporal intermittency which occurs here seems to be unrelated to the spacial intermittency discussed elsewhere.

oscillations (Ruelle [63]).[8] This *chemical turbulence* is indeed seen experimentally, and has provided the first example of a strange attractor reconstructed from experimental data (see Roux, Rossi, Bachelart and Vidal [62]). All kinds of electromechanical systems also exhibit turbulent time evolution, described as *deterministic noise*, which can now be correctly identified and studied, and may play a significant practical role.

Sensitive dependence on initial condition for dissipative systems has come to be called *chaos*. The chief requirement for a nonlinear system to exhibit chaos is that its phase space have at least 3 dimensions.

Strange attractors and chaotic behavior should also occur in biological systems. The case of ecological models has been discussed by May [47]. Aperiodic cyclès are also expected in macroeconomics [69]. Since the experimental conditions in ecology and economics cannot be precisely controlled, precise predictions are also not expected, but it is at least of philosophical interest to perceive the dynamical causes of chaos in these disciplines.

Sensitive dependence on initial condition

By differentiating (3) we obtain the evolution equation for tangent vectors:

$$\frac{dW}{dt} = (D_{w(t)}F_{\mu})W, \tag{5}$$

where D denotes the derivative. Sensitive dependence on initial condition arises if W grows exponentially with time.

It is part of popular wisdom that, for certain particular initial conditions w_0 of the time evolutions which occur in nature, a small change may lead after a while to very different situations. (A pebble at the top of a mountain may fall on one side or the other). It is less obvious that for some dynamical systems there is an exponentially growing W for every initial w_0. That this is so for the geodesic flow on a surface of negative curvature was shown by Hadamard [21]. There are thus natural systems for which no precise prediction can be made, because any small imprecision on the initial condition will result in a large uncertainty for the future behavior. In 1906, P. Duhem [11], referring to Hadamard's work, stressed the philosophical importance of this fact for the problem of predictability

[8] While such a prediction may seem trivial in retrospect, things did not appear so at the time, and reference [63] shared with [75] the fate of not being accepted by the first journal to which it was submitted for publication. Experimentalists are, by the way, absolutely right in treating "new theoretical ideas" with great caution.

in physics.[9] In 1908, H. Poincaré [59] also emphasises the importance of sensitivity to initial condition in a discussion of chance. He already considers meteorology and recognizes the reason why weather predictions are imprecise (see [59], p. 69). He also gives the example of a gas, where a little change in the initial data for one molecule will be amplified by collisions until a molecule which should have hit another one now misses it, so that the microscopic dynamics of the gas has now become completely different. M. Berry[10] has estimated that it would take only 50 collision times to reach this result, if the initial perturbation was the gravitational action of one electron located at the limit of the known universe!

Our understanding of developed turbulence is quite imperfect, but nevertheless gives us the possibility to estimate the growth of fluctuations in a system like the earth's atmosphere. This problem has been studied by Lorenz, Kraichnan and Leith, relevant times are of the order of a week or two. In the kind of turbulence present in air above a radiator it may take of the order of one minute for molecular fluctuations to be amplified to the macroscopic level. (This estimate uses Kolmogorov's model of turbulence, see Ruelle [68]). Putting these facts together the reader is left to imagine how the gravitational effect of an electron at the edge of the universe may affect his fate and change the course of his life. Even if we suppose that the deterministic laws of classical mechanics govern the evolution of our universe, we see that the introduction of chance and probabilities is a necessity in the practice of physics and in everyday life.

It may occur to the reader that, among other things, the position of the planets in the sky may influence his life. Does this justify the claims of astrology? On the contrary, the arguments which show how the planets may influence human fate also indicate that such influences are, for all practical purposes, impredictable.

Ergodic theory of differentiable dynamical systems

Although Hadamard, Duhem and Poincaré understood the origins and implications of sensitive dependence on initial conditions, a quantitative treatment of this notion came much later. The relevant concepts belong to ergodic theory, they are the *entropy* (invariant of Kolmogorov [28] and Sinai [78]) and the *characteristic exponents* (or "Liapounov" exponents, defined in general by Oseledec [54] only in 1968). A dynamical system

[9] The relevant section of Duhem's book is entitled "Exemple de déduction mathématique à tout jamais inutilisable". It was kindly pointed out to me by R. Thom.

[10] Private communication.

known with finite precision acts as a random number generator, and the entropy describes the rate of information production by the system. The characteristic exponents describe the rate of increase of the perturbations of the initial condition. The time rates involved in these definitions are defined with respect to a measure ϱ invariant under time evolution (ϱ is the ergodic measure corresponding to time averages). The brilliant work of N. S. Krylov [31] anticipates the notions of entropy and characteristic exponents, but also confuses them as noted by Sinai.

The idea to study differentiable dynamical systems almost everywhere with respect to invariant measures was developed by Pesin [55], [56], [57] (and later Katok [25], etc.) and turned out to be remarkably fruitful. In particular, the construction of stable and unstable manifolds almost everywhere can be extended to the infinite dimensional situation of hydrodynamics (Ruelle [67], [73], Mañé [46]).

An important conceptual problem arises now: what are the invariant measures μ which describe turbulence? In classical mechanics there is a natural measure because of unique ergodicity. But a strange Axiom A attractor has uncountably many different ergodic measures. Which one is physically relevant? I.e., which one reproduces time averages? Perhaps the one which maximizes entropy? This guess is *wrong*! It is natural to look for measures which are invariant under small stochastic perturbations (Kolmogorov) but this does not solve the problem. A satisfactory answer has been obtained first in the Anosov case (Sinai [79]), then in the general Axiom A case (Ruelle [64], Bowen and Ruelle [4], Kifer [26]). There, the time averages for almost all initial conditions with respect to Lebesgue measure in the neighborhood of an attractor yields the same measure μ on the attractor, and this measure is stable under small stochastic perturbations. The measure μ has conditional measures on unstable manifolds which are absolutely continuous with respect to Lebesgue measure, and its entropy is the sum of the positive characteristic exponents. I believe that this situation has some generality (see Ruelle [65], [66]). At present there are both counterexamples (Bowen, Katok [25]) and deep positive results (Pugh and Shub [61], Ledrappier [37], [38]). General results on the Hausdorff dimension of attractors are also known (Mañé [45], Douady et Oesterlé [10], Ledrappier [36], L.-S. Young [83]).

Developed turbulence

For completeness we deal here very briefly with the vast subject of developed turbulence. Of central importance is the Kolmogorov [27] theory (discovered by Kolmogorov, Oboukhov, Onsager, etc.). This physical theory describes the energy cascade from large spatial structures (where energy is injected) to small "eddies" (where energy is dissipated by viscosity). As noted by Landau [34], Kolmogorov theory is essentially a consequence of arguments of dimensional analysis. This explains why it is robust and successful. However, this theory does not take into account the important phenomenon of intermittency, i.e., the fact that most of the vorticity and dissipation is found in a small subset of physical space. Intermittency in 2 dimensions was noted by Poincaré [58] (existence of localized vortices, visible at the surface of a river for instance). Poincaré tried to give an explanation based on stability of the motion arguments. Onsager [53] proposed an explanation based on the statistical mechanics of vortices, and using negative temperatures. This explanation has been challenged recently (Fröhlich and Ruelle [18]). Three-dimensional intermittency can be modelled (see Frisch, Sulem and Nelkin [17] and references given there) and numerical experiments give a coherent picture (see in particular Chorin [8]) of a self-similar structure (see Mandelbrot [44]). The associated self-similarity dimension is numerically ≈ 2.6, which is much larger than the dimension of the set of possible singularities of solutions of the Navier–Stokes equation. (The existence of such singularities is thus apparently unrelated to intermittency). Elements of an ergodic analysis of Navier–Stokes time evolution have been given (Ruelle [74]). However, a deductive theory of developed turbulence does not exist, and a mathematical basis for the important theoretical literature on this subject is still lacking. (See the monographs of Monin and Yaglom [51], Batchelor [2], the papers of Kraichnan [29], [30], etc.).

Conclusion

The application of non trivial mathematical ideas has given us some understanding of the onset of hydrodynamic turbulence, and changed our notions on turbulence in general. It is good to assess — without epistemological prejudice — the articulation of mathematics and physics in this example. First it must be admitted that the success obtained here did not depend on the proof of some very difficult and deep theorem (even though non trivial theorems have been proved in the course of the study). Does this mean that the important new fact was the revelation of a general philosophical principle like sensitive dependence on initial condition? In fact no: Duhem and Poincaré had a perfectly clear understanding of the principle of sensitive dependence on initial condition, and of its consequences. But the principle was not embodied in mathematical theories of turbulence.

If one looks at the conceptual framework of the modern theory of the onset of turbulence, one finds mathematically sophisticated objects like strange attractors, characteristic exponents, or the Feigenbaum bifurcation. It is important that these objects become useful before their mathematical status is completely elucidated. In other words our understanding of phenomena is made possible by the use of pieces of sophisticated mathematical reasoning, connecting experimental data, computer evidence, and physical assumptions. A purely deductive analysis starting with the Navier–Stokes equation is not attempted: it does not appear feasible at this moment, and might be inappropriate because of the approximate nature of the Navier–Stokes equation.

Looking back at the history of science, we would see that the kind of close interaction which we find here between mathematical ideas and experimental facts has been a regular feature of the evolution of physics. This interaction is of considerable mathematical and philosophical value: it gives us a glimpse, different from intuition, into mathematical reality which is not yet mathematical theory.

References

[1] Ahlers G., Low-Temperature Studies of the Rayleigh-Bénard Instability and Turbulence, *Phys. Rev. Lett.* **33** (1974), pp. 1185–1188.

[2] Arnold V., Sur la géométrie différenteille des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, *Ann. Inst. Fourier* **16** (1966), pp. 319–361.

[3] Batchelor G. K., *The Theory of Homogeneous Turbulence*, University Press, Cambridge, 1967.

[4] Bowen R., and Ruelle D., The Ergodic Theory of Axiom A Flows, *Inventiones math.* **29** (1975), pp. 181–202.

[5] Caffarelli L., Kohn R. and Nirenberg L., *Partial Regularity of Suitable Weak Solutions of the Navier–Stokes Equations*, Preprint.

[6] Campanino M. and Epstein H., On the Existence of Feigenbaum's Fixed Point, *Commun. Math. Phys.* **79** (1981), pp. 261–302.

[7] Campanino M., Epstein H. and Ruelle D., On Feigenbaum's Functional Equation $g\circ g(\lambda x)+\lambda g(x) = 0$, *Topology* **21** (1982), pp. 125–129.

[8] Chorin A. J., The Evolution of a Turbulent Vortex, *Commun. Math. Phys.* **83** (1982), pp. 517–535.

[9] Collet P., Eckmann J.-P. and Koch H., Period Doubling Bifurcations for Families of Maps on $\boldsymbol{R}^n$, *J. Statist. Phys.* **25** (1981), pp. 1–14.

[10] Douady A. and Oesterlé J., Dimension de Hausdorff des attracteurs, *C. R. Acad. Sc. Paris* **290** A (1980), pp. 1135–1138.

[11] Duhem P., *La théorie physique. Son objet et sa structure*, Chevalier et Rivière, Paris, 1906.

[12] Eckmann J.-P., Roads to Turbulence in Dissipative Dynamical Systems, *Rev. Mod. Phys.* **53** (1981), pp. 643–654.
[13] Feigenbaum M. J., Quantitative Universality for a Class of Nonlinear Transformations, *J. Statist. Phys.* **19** (1978), pp. 25–52.
[14] Feigenbaum M. J., The Universal Metric Properties of Nonlinear Transformations, *J. Statist. Phys.* **21** (1979), pp. 669–706.
[15] Feigenbaum M. J., The Transition to Aperiodic Behavior in Turbulent Systems, *Commun. Math. Phys.* **77** (1980), pp. 65–86.
[16] Foias C. and Temam R., Some Analytic and Geometric Properties of the Solutions of the Evolution Navier–Stokes Equations, *J. Math. pures et appl.* **58** (1979), pp. 339–368.
[17] Frisch U., Sulem P.-L. and Nelkin M., A Simple Dynamical Model of Intermittent Fully Developed Turbulence. *J. Fluid Mech.* **87** (1978), pp. 719–736.
[18] Fröhlich J. and Ruelle D., Statistical Mechanics of Vortices in an Inviscid Two-Dimensional Fluid, *Commun. Math. Phys.* **87** (1982), pp. 1–36.
[19] Gollub J. P. and Swinney H. L., Onset of Turbulence in a Rotating Fluid, *Phys. Rev. Lett.* **35** (1975), pp. 927–930.
[20] Guckenheimer J., A Strange, Strange Attractor, In: *The Hopf bifurcation theorem and its applications*, J. E. Marsden and M. McCracken (ed.), Springer, New York, 1976, pp. 368–381.
[21] Hadamard J., Les surfaces à courbures opposées et leurs lignes géodésiques *J. Math. pures et appl.* **4** (1898), pp. 27–73.
[22] Hopf E., Abzweigung einer periodischen Lösung von einer stationären Lösung eines Differentialsystems, *Ber. Math.-Phys. Kl. Sächs. Akad. Wiss. Leipzig* **94** (1942), pp. 1–22.
[23] Hopf E., A Mathematical Example Displaying the Features of Turbulence, *Commun. pure appl. Math.* **1** (1948), pp. 303–322.
[24] Hopf E., Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen, *Math. Nachrichten* **4** (1951), pp. 213–331.
[25] Katok A., Lyapunov Exponents, Entropy and Periodic Orbits for Diffeomorphisms, *Publ. math. IHES* **51** (1980), pp. 137–174.
[26] Kifer Yu. I., On Small Random Perturbations of Some Smooth Dynamical Systems, *Izv. Akad. Nauk SSSR, Ser. Mat.* **38**, № 5 (1974), pp. 1091–1115. English translation: *Math. USSR Izvestija* **8** (1974), pp. 1083–1107.
[27] Kolmogorov A. N., Local Structure of Turbulence in an Incompressible Fluid at Very High Reynolds Numbers, *Dokl. Akad. Nauk SSSR* **30** (1941), pp. 299–303.
[28] Kolmogorov A. N., A New Metric Invariant of Transient Dynamical Systems and Automorphisms of Lebesgue Spaces, *Dokl. Akad. Nauk SSSR* **119** (1958), pp. 861–864.
[29] Kraichnan R. H., Inertial Ranges in Two-Dimensional Turbulence, *Phys. Fluids* **10** (1967), pp. 1417–1423.
[30] Kraichnan R. H., On Kolmogorov's Inertial-Range Theories, *J. Fluid Mech.* **62** (1974), pp. 305–330.
[31] Krylov N. S., *Works on the Foundations of Statistical Physics*, Princeton University Press, Princeton NJ, 1979.
[32] Ladyzhenskaya O. A., *The Mathematical Theory of Viscous Incompressible Flow*, 2nd ed., Nauka, Moscow, 1970; 2nd English ed., Gordon and Breach, New York, 1969.

[33] Landau L. D., On the Problem of Turbulence, *Dokl. Akad. Nauk SSSR* **44**, № 8 (1944), pp. 339–342.
[34] Landau L. D. and Lifshitz E. M., *Fluid Mechanics*, Pergamon Press, Oxford, 1959.
[35] Lanford O. E., A Computer-Assisted Proof of the Feigenbaum Conjectures, *Bull. Amer. Math. Soc.* **6** (1982), pp. 427–434.
[36] Ledrappier F., Some Relations Between Dimension and Lyapounov Exponents, *Commun. Math. Phys.* **81** (1981), pp. 229–238.
[37] Ledrappier F., Some Properties of Absolutely Continuous Invariant Measures on an Interval, *Ergod. Th. and Dynam. Syst.* **1** (1981), pp. 77–93.
[38] Ledrappier F., *Propriétés ergodiques des mesures de Sinaï*, Preprint.
[39] Leray J., Sur le mouvement d'un liquide visqueux emplissant l'espace, *Acta Math.* **63** (1934), pp. 193–248.
[40] Libchaber A. and Maurer J., Une expérience de Rayleigh–Bénard de géométrie réduite; multiplication, accrochage et démultiplication de fréquences, *J. Physique* **41**, Colloques C3, (1980), pp. 51–56.
[41] Lions J. L., *Quelques méthodes de résolution des problèmes aux limites non-linéaires*, Dunod, Paris, 1969.
[42] Lorenz E. N., Deterministic Non Periodic Flow, *J. atmos. Sci.* **20** (1963), pp. 130–141.
[43] Mallet-Paret J., Negatively Invariant Sets of Compact Maps and an Extension of a Theorem of Cartwright, *J. Diff. Equ.* **22**, (1976), pp. 331–348.
[44] Mandelbrot B., *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
[45] Mañé R., On the Dimension of the Compact Invariant Sets of Certain Nonlinear Maps. In: *Dynamical systems and turbulence, Warwick 1980*, Lecture Notes in Math. 898, Springer, New-York, 1981, pp. 230–242.
[46] Mañé R., *Lyapunov Exponents and Stable Manifolds for Compact Transformations*, Preprint.
[47] May R. M., Simple Mathematical Models with Very Complicated Dynamics, *Nature* **261** (1976), pp. 459–467.
[48] McLaughlin J. B. and Martin P. C., Transition to Turbulence of a Statically Stressed Fluid, *Phys. Rev. Lett.* **33** (1974), pp. 1189–1192.
[49] McLaughlin J. B. and Martin P. C., Transition to Turbulence in a Statically Stressed Fluid System, *Phys. Rev.* A **12** (1975), pp. 186–203.
[50] Monin A. S., On the Nature of Turbulence, *Uspekhi Fiz. Nauk* **125**, 1 (1978), pp. 97–122. English translation: *Soviet Physics Uspekhi* **21** (1978), pp. 429–442.
[51] Monin A. S. and Yaglom A. M., *Statistical Fluid Mechanics: Mechanics of turbulence*, Nauka, Moscow, 1965. English translation: 2 vols (edited by J. L. Lumley), MIT Press, Cambridge, Mass, 1971 and 1975.
[52] Newhouse S., Ruelle D. and Takens F., Occurrence of Strange Axiom *A* Attractors Near Quasiperiodic Flows on T^m, $m \geqslant 3$. *Commun. Math. Phys.* **64** (1978), pp. 35–40.
[53] Onsager L., Statistical Hydrodynamics, *Suppl. Nuovo Cim.* **6** (1949), pp. 279–287.
[54] Oseledec V. I., A Multiplicative Ergodic Theorem. Lyapunov Characteristic Numbers for Dynamical Systems, *Trudy Moskov. Mat.* Obšč. **19** (1968), pp. 179–210. English translation: *Trans. Moscow Math. Soc.* **19** (1968), pp. 197–221.

[55] Pesin Ya. B., Lyapunov Characteristic Exponents and Ergodic Properties of Smooth Dynamical Systems With an Invariant Measure, *Dokl. Akad. Nauk SSSR* **226**, 4 (1976), pp. 774–777. English translation: *Soviet Math. Dokl.* **17**, 1 (1976), pp. 196–199.

[56] Pesin Ya. B. Invariant Manifold Families Which Correspond to Nonvanishing Characteristic Exponents. *Izv. Akad. Nauk SSSR Ser. Mat.* **40**, 6 (1976), 1332–1379. English translation: *Math. USSR Izv.* **10**, 6 (1976), pp. 1261–1305.

[57] Pesin Ya. B., Lyapunov Characteristic Exponents and Smooth Ergodic Theory, *Uspehi Mat. Nauk* **32**, 4 (196), (1977), pp. 55–112. English translation: *Russian Math. Surveys* **32**, 4 (1977), pp. 55–114.

[58] Poincaré H., *Théorie des tourbillons*, Georges Carré, Paris, 1893.

[59] Poincaré H., *Science et méthode*, Ernest Flammarion, Paris, 1908.

[60] Pomeau Y. and Manneville P., Intermittent transition to turbulence in dissipative dynamical systems. *Commun. math. Phys.* **74** (1980), pp. 189–197.

[61] Pugh C. C. and Shub M., Differentiability and Continuity of Invariant Manifolds, *Ann. NY Acad. Sci.* **357** (1980), pp. 322–239.

[62] Roux J. C., Rossi A., Bachelart S. and Vidal C., Representation of a Strange Attractor From an Experimental Study of Chemical Turbulence, *Phys. Letters* **77** A (1980), pp. 391–393.

[63] Ruelle D., Some Comments on Chemical oscillations, *Trans. N. Y. Acad. Sci.* II, **35** (1973), pp. 66–71.

[64] Ruelle D., A Measure Associated with Axiom A Attractors, *Amer. J. Math.* **98** (1976), pp. 619–654.

[65] Ruelle D., Dynamical Systems with Turbulent Behavior. In: *Mathematical problems in theoretical Physics*, Lecture Notes in Physics 80, Springer, Berlin, 1978, pp. 341–360.

[66] Ruelle D., What are the Measures Describing Turbulence? *Progr. Theoret. Phys. Suppl.* **64** (1978), pp. 339–345.

[67] Ruelle D., Ergodic Theory of Differentiable Dynamical Systems, *Publ. math. IHES* **50** (1979), pp. 275–306.

[68] Ruelle D., Microscopic Fluctuations and Turbulence, *Phys. Letters* **72** A (1979), pp. 81–82.

[69] Ruelle D., Strange Attractors. *Math. Intelligencer* **2** (1980), pp. 126–137.

[70] Ruelle D., Measures Describing a Turbulent Flow, *Ann. N. Y. Acad. Sci.* **357** (1980), pp. 1–9.

[71] Ruelle D., Differentiable Dynamical Systems and the Problem of Turbulence, *Bull. Amer. Math. Soc.* **5** (1981), pp. 29–42.

[72] Ruelle D., Small Random Perturbations of Dynamical Systems and the Definition of Attractors, *Commun. math. Phys.* **82** (1981), pp. 137–151.

[73] Ruelle D., Characteristic Exponents and Invariant Manifolds in Hilbert Space, *Annals of Math.* **115** (1982), pp. 243–290.

[74] Ruelle D., Large Volume Limit of the Distribution of Characteristic Exponents in Turbulence, *Commun. math. Phys.* **87** (1982), pp. 287–302.

[75] Ruelle D. and Takens F., On the Nature of Turbulence, *Commun. math. Phys.* **20**, (1971), pp. 167–192. Note concerning our paper On the Nature of Turbulence, *Commun. math. Phys.* **23** (1971), pp. 343–344.

[76] Scheffer V., The Navier–Stokes Equations on a Bounded Domain, *Commun. math. Phys.* **73** (1980), pp. 1–42.

[77] Serrin J., The Initial Value Problem for the Navier–Stokes Equations, In: *Nonlinear Problems*, R. F. Langer (ed.), University of Wisconsin Press, 1963, pp. 69–98.
[78] Sinai Ia. G., On the Concept of Entropy of a Dynamical System, *Dokl. Akad. Nauk SSSR* **124** (1959), pp. 768–771.
[79] Sinai Ia. G., Gibbs Measures in Ergodic Theory, *Uspekhi Mat. Nauk* **27**, 4 (1972), pp. 21–64. English translation: *Russian Math. Surveys* **27**, 4 (1972), pp. 21–69.
[80] Smale S., Differentiable Dynamical Systems, *Bull. Amer. Math. Soc.* **73** (1967), pp. 747–817.
[81] Temam R., *Navier–Stokes Equations*, revised edition, North-Holland, Amsterdam, 1979.
[82] von Neumann J., Recent Theories of Turbulence (1949), *Collected Works*, 6 Pergamon, Oxford, 1963, pp. 437–472.
[83] Young L.-S., Capacity of Attractors, *Ergod. Th. and Dynam. Syst.* **1** (1981), pp. 381–388.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

YUM-TONG SIU

Some Recent Developments in Complex Differential Geometry

We survey here some recent results in the theory of complex manifolds obtained by the methods of global differential geometry. These results stemmed from attempts to generalize to the higher-dimensional case the uniformization theorem for Riemann surfaces. A simply connected Riemann surface must be either the Euclidean plane, the Riemann sphere, or the open unit disc in the Euclidean plane. In the higher-dimensional case there is no such simple classification of simply connected complex manifolds, even with the imposition of very strong curvature conditions. A slight perturbation of the open unit ball in C^n for $n > 1$ will in general give us a new complex structure (not biholomorphic to the ball) [9], where the Bergman metric is still complete and has sectional curvature very close to that of the ball [16]. We will discuss the various results in higher dimensions for the parabolic, elliptic, and hyperbolic cases, generalizing those for Riemann surfaces. At the end we mention some recent results on K3 surfaces.

§ 1. Parabolic case

The problem concerns the characterization of C^n as a Kähler manifold with sufficiently fast decay of curvature. The first result of this kind was obtained by Siu and Yau [55] in response to a question posed by Greene and Wu [18]: A simply connected complete Kähler manifold of complex dimension $n \geqslant 2$ with $0 \geqslant$ sectional curvature $\geqslant -A/(1+r^{2+\varepsilon})$ (where A, $\varepsilon > 0$ and r is the distance from a fixed point) is biholomorphic to C^n. The method is to use the L^2 estimates of $\bar\partial$ to produce n holomorphic functions of nearly linear growth which are local coordinates at one point. The map from the manifold to C^n defined by these n functions is a local biholomorphism, because the growth of the Jacobian n-form is related to the growth of the volume of its zero-set. Moreover, the map is proper, because the manifold is Stein and the power series obtained by expanding any global holomorphic function in terms of these n functions at any point converges uniformly on any compact subset of the manifold. The curvature condition is used to get suitable weight functions for the L^2 estimates and to get sup norm estimates for holomorphic functions and n-forms with weighted L^2 bounds. Later Greene and Wu [19] made some quantitative improvements on the assumption of the decay order of the curvature.

Recently Mok, Siu and Yau [36] showed that under the above assumption the manifold is actually biholomorphically isometric to C^n. Moreover, if the curvature condition is replaced by $|\text{sectional curvature}| \leqslant A_\varepsilon/(1+r^{2+\varepsilon})$ (where A_ε is a positive number depending on ε) and the exponential map at the fixed point is a diffeomorphism, then the manifold is still biholomorphic to C^n. In the case of positive curvature of fast decay, if the scalar curvature is $\leqslant A/(1+r^{2+\varepsilon})$ and the volume of the geodesic ball centered at the fixed point of radius s is $\geqslant Cs^{2n}$ (C being a positive constant), then the manifold is biholomorphically isometric to C^n when the sectional curvature is non-negative or when the bisectional curvature is non-negative and the manifold is Stein. These two isometry results are obtained as follows. One solves the equation $\partial\bar{\partial}u = \text{Ricci}$ either by the L^2 estimates of $\bar{\partial}$ or by solving first $\Delta u = \text{scalar curvature}$ and applying the Bochner technique to $\Delta\,|\partial\bar{\partial}u - \text{Ricci}|^2$ and then one proves that u must be constant.

The isometry results in Mok, Siu and Yau [36] prompted the conjecture for similar results in Riemannian geometry. Such results in Riemannian geometry were afterwards obtained by Greene and Wu [20] and also independently by Gromov [21].

An open problem about the characterization of C^n is whether a complete non-compact Kähler manifold with positive sectional curvature is biholomorphic to C^n. The case $n = 2$ under some mild additional conditions was recently proved by Mok [35] together with some other partial results. Greene and Wu [17] showed that such a manifold is Stein. (Under the weaker assumption of positive bisectional curvature such a manifold is conjectured to be Stein. Though it is easy to produce global holomorphic functions separating points and giving local coordinates, the conjecture remains unproved.)

There is another kind of characterization of C^n. It is by the existence of an exhaustion function satisfying a certain Monge–Ampère equation. This characterization works not only for C^n but also for bounded balls in C^n. So it is also a result in the hyperbolic case. Suppose M is an n-dimensional complex manifold with a strongly plurisubharmonic exhaustion function $u\colon M \to [0, r^2)$ $(0 < r \leqslant \infty)$ such that $(\partial\bar\partial \log u)^n = 0$, $\sqrt{-1}\,\partial\bar\partial\log u \geqslant 0$, and $(\partial\bar\partial\log u)^{n-1} \neq 0$. Stoll [59] proved that M with the Kähler metric whose potential is u is biholomorphically isometric to the ball of radius r in $\boldsymbol{C}^n$ with the metric induced from the Euclidean metric of $\boldsymbol{C}^n$. Alternative proofs and some generalizations were given by Burns [7]. Another alternative proof was given by P.-M. Wong [65]. The idea is to consider complex curves in M whose tangent spaces are the annihilators of $\partial\bar\partial\log u$ and to show that the space of all such curves is $\boldsymbol{P}_{n-1}$. One then establishes the isometry by using the tautological line bundle of $\boldsymbol{P}_{n-1}$.

§ 2. Elliptic case

The problem concerns the curvature characterization of the complex projective space or, more generally, irreducible compact Hermitian symmetric manifolds. Frankel [14] conjectured that a compact n-dimensional Kähler manifold with positive sectional curvature is biholomorphic to $\boldsymbol{P}_n$. He and Andreotti [14] gave a proof for $n = 2$. Mabuchi [33] proved it for $n = 3$. Mori [37], using the methods of algebraic geometry of characteristic $p > 0$, proved the general case with the weakened assumption that the tangent bundle is ample. Siu and Yau [56] gave a differential-geometric proof for the general case with the assumption that the bisectional curvature is positive.

The differential-geometric proof of the Frankel conjecture makes use of harmonic maps. The energy $E(f)$ of a map $f\colon N \to M$ between two Riemannian manifolds is the L^2 norm over N of its differential df. The map f is harmonic if it is critical for the energy functional. In particular, it is harmonic if it minimizes energy among all maps in its homotopy class. Hartshorne and Kobayashi and Ochiai [28, 29] observed that a compact Kähler manifold M with positive bisectional curvature is biholomorphic to $\boldsymbol{P}_n$ if a generator of $H^2(M, \boldsymbol{Z})$ can be represented by a rational curve. On the other hand, by the result of Sachs and Uhlenbeck [44], one can represent in this case a generator of $H^2(M, \boldsymbol{Z})$ by a finite sum of harmonic maps $f_i\colon \boldsymbol{P}_1 \to M$ so that $\sum E(f_i)$ is minimum among maps in the homotopy class of $\sum f_i$. By using the second variation of the energy functional, one concludes from the positivity of the bisectional curvature that each f_i is either holomorphic or antiholomorphic (see also [47]). There can be only one f_i, otherwise we have one holomorphic f_i and one antiholomorphic f_j and, by the positivity of the tangent bundle of M, the images of f_i and

f_j can be holomorphically deformed until they touch at some point and one can then modify f_i+f_j near the point of contact to produce a map in the homotopy class of f_i+f_j with smaller energy. Thus a generator of $H^2(M, \mathbf{Z})$ is represented by a rational curve.

The above method was refined in Siu [49] to yield the following curvature characterization of the complex hyperquadric. A compact Kähler manifold M of complex dimension $\geqslant 3$ with non-negative bisectional curvature is biholomorphic to either the complex projective space or the complex hyperquadric if at every point there exists no tangent subspace V of complex dimension 2 such that the bisectional curvature in the direction of any two vectors always vanishes when one of them is in V.

The general problem is to determine whether a compact Kähler manifold with non-negative sectional (or bisectional) curvature and positive definite Ricci curvature is necessarily biholomorphic to a compact Hermitian symmetric manifold. Moreover, if it is biholomorphic to an irreducible compact Hermitian symmetric manifold of rank at least two, one wants to know whether the Kähler metric with non-negative sectional (or bisectional) curvature is necessarily an invariant metric.

In Siu [52] the following weak partial result on the above problem was obtained by using the Bochner–Kodaira technique and holonomy groups. Let M be a compact Kähler manifold whose cotangent bundle with the induced metric is seminegative in the sense of Nakano [41]. Assume that at some point of M it is not possible to decompose the tangent space of M into two non-zero orthogonal direct summands such that the bisectional curvature in the direction of two vectors with one in each summand must vanish. Then either M is an irreducible Hermitian symmetric manifold with respect to the given Kähler metric or its cohomology ring with real coefficients is isomorphic to that of the complex projective space of the same dimension. This partial result is unsatisfactory, because seminegativity in the sense of Nakano is a curvature operator condition.

Recently Bando [4] proved that a three-dimensional compact Kähler manifold of non-negative bisectional curvature whose Ricci curvature is semipositive everywhere and positive definite somewhere is biholomorphic to the complex hyperquadric or a product of complex projective spaces. His proof makes use of the methods of Siu and Yau [56], Siu [49], the classification of compact Kähler surfaces of non-negative bisectional curvature by Howard and Smyth [25], the work of Howard, Smyth and Wu [26] and Wu [67] on the splitting of Kähler manifolds of non-negative bisectional curvature, and Hamilton's method [22] of solving evolution equations to increase the positivity of curvature.

For this problem, if the Kähler metric is assumed to be Einstein, then the result of Berger [5] and Gray [15] tells us that a Kähler–Einstein metric on a compact complex manifold is locally symmetric if its sectional curvature is non-negative. It is unknown in dimension at least four whether in this result one can replace sectional curvature by bisectional curvature.

§ 3. Hyperbolic case

The problem is to produce non-constant bounded holomorphic functions on a simply connected complete Kähler manifold whose sectional (or bisectional) curvature is bounded from above by a negative number, with the goal of embedding it into a bounded domain. Very little is known about this problem. Wu [66] observed that in the case of negative sectional curvature the manifold is Stein. It is an open problem whether the manifold is Stein in the case of negative bisectional curvature. It is even unknown whether such a manifold is necessarily non-compact. Recently B. Wong [64] constructed a simply connected compact complex surface with ample cotangent bundle. This would be a negative answer to the last question about non-compactness if the ampleness of the cotangent bundle could be replaced by the negativity of the bisectional curvature of some Kähler metric. On the other hand, in the corresponding situation in Riemannian geometry, Anderson [1] and Sullivan [61] recently succeeded in constructing non-constant bounded harmonic functions on any simply connected complete Riemannian manifold whose sectional curvature is bounded from above by a negative number. (Earlier Prat [*C.R. Acad. Sci. Paris* **284** (1977), pp. 687–690] did the real 2-dimensional case under additional conditions.) Later a very simple construction was given by Schoen [46].

Even when one confines oneself to the special case of manifolds which are the universal covers of compact Kähler manifolds with negative sectional (or bisectional) curvature, one still does not know how to construct non-constant bounded holomorphic functions. For a long time, for lack of examples, it was believed that every compact Kähler manifold with negative sectional curvature is covered by the ball. This belief was reinforced by the result of B. Wong [63] characterizing balls as smooth bounded (strongly pseudoconvex) domains admitting compact quotients and by Yang's result [68] that a bidisc cannot cover a compact Kähler surface with negative bisectional curvature. Finally an example was constructed by Mostow and Siu [40] of a compact Kähler surface with negative sectional curvature not covered by the ball. Actually its curvature is even strongly negative in the sense of [48]. The manifold is constructed by using an almost discrete subgroup of the automorphism group of the ball generated by three complex reflections [39] and the metric is obtained by piecing together the Bergman metrics of the ball and a domain ramified over the ball along a complex line. For this construction one can also use the almost discrete subgroup constructed from the extension by Deligne and Mostow [11] of a method of Picard [43].

It is not known whether requiring the Kähler metric to be Einstein would force the universal cover of a negatively curved compact Kähler manifold to be the ball. A partial result in the surface case was obtained by Siu and Yang [54]. This question is related to the problem of obtaining conditions under which the Kähler–Einstein metric constructed by Yau [69] has negative sectional curvature.

Compact Kähler manifolds of complex dimension at least two with suitable negative curvature conditions enjoy the property of strong rigidity. In the Riemannian case Mostow [38] obtained the following result on strong rigidity. Two compact locally symmetric Riemannian manifolds of non-positive sectional curvature are isometric (up to normalizing constants) if they are of the same homotopy type and they do not admit closed geodesic submanifolds of real dimension $\leqslant 2$ which are locally direct factors. In the complex case one has a stronger kind of strong rigidity. A compact Kähler manifold is said to be strongly rigid if any compact Kähler manifold of the same homotopy type is either biholomorphic or antibiholomorphic to it. Siu [48] proved that a compact Kähler manifold of complex dimension at least two is strongly rigid if its Kähler metric makes its cotangent bundle positive in the sense of Nakano (or if the weaker condition of strongly negative curvature in the sense of [48] is satisfied). The idea of the proof is as follows. Let M be the compact Kähler manifold with strongly negative curavture and let N be a compact Kähler manifold of the same homotopy type as M. By Eells and Sampson [13], there exists a harmonic map f: $N \to M$ which is a homotopy equivalence. By applying the Bochner–Kodaira technique to $\bar{\partial} f$, one obtains the vanishing of a term involving $\bar{\partial} f$, ∂f, and the curvature of M. The curvature condition on M forces the vanishing of $\bar{\partial} f$ or ∂f at points where the rank of df over $\boldsymbol{R}$ is at least 4. The proof can be regarded as a quasilinear analog of the vanishing theorem of Kodaira for negative bundles in which $\bar{\partial} f$ takes the place of a harmonic form.

This method of proof can be used to obtain the holomorphicity of harmonic maps and strong rigidity in more general cases [52]. Let M be a compact Kähler manifold whose cotangent bundle with the induced metric is non-negative in the sense of Nakano. Suppose for some positive integer p the bundle of $(p, 0)$-forms on M with the induced metric is positive in the sense of Nakano and suppose at every point of M it is not possible to find two non-zero complex tangent subspaces V, W such that the sum of the complex dimensions of V and W exceeds p and the bisectional curvature in the direction of one vector of V and one vector of W always vanishes. Then any harmonic map f from a compact Kähler manifold to M must be either holomorphic or antiholomorphic if the rank of the differential df of f over $\boldsymbol{R}$ is at least $2p+1$ at some point. In particular, M is strongly rigid if the complex dimension of M exceeds p. As a consequence, one concludes that a compact quotient of an irreducible bounded symmetric domain of complex dimension at least two is strongly rigid [48, 50, 52]. For a compact quotient of an irreducible bounded symmetric domain Siu [50, 52] and Zhong [70], by using the results of Calabi and Vesentini [10] and Borel [6], computed the smallest p for which the above conditions hold.

The above result on the holomorphicity of harmonic maps enables one to construct complex submanifolds in a compact Kähler manifold with suitable negative curvature conditions and also it leads to results on the relationship between the Kuranishi deformation space [30] and the Douady deformation space [12] for a submanifold of a manifold with suitable negative curvature conditions (see Kalka [27] and Siu [52]).

It is possible to obtain strong rigidity results in the case of non-compact manifolds of finite volume (Siu and Yau [58]). Morevoer, Siu and Yau [57] proved that a complete Kähler manifold of finite volume whose sectional curvature is bounded between two negative numbers can be compactified by adding a finite number of points so that the compactification can be blown up at the added points to become a projective algebraic variety. This generalizes the result on the corresponding locally symmetric cases by Satake [45] and Baily and Borel [3] (see also Andreotti and Grauert [2]).

§ 4. K3 surfaces

One of the most interesting applications of the methods of global differential geometry to the theory of complex manifolds are the results recently obtained on K3 surfaces. A K3 surface is a simply connected compact complex surface with trivial canonical line bundle. Let X be a smooth manifold diffeomorphic to one (and hence every) K3 surface. Let V be $H^2(X, \boldsymbol{Z})$ endowed with the quadratic form defined by the cup product and let Ω be the set of all real 2-planes in $V \otimes \boldsymbol{R}$ on which the quadratic form is positive definite. A marking on a K3 surface M is an isomorphism between $H^2(M, \boldsymbol{Z})$ and V compatible with the cup product. The period of a marked K3 surface is the element of Ω spanned by the elements of $H^2(M, \boldsymbol{R})$ corresponding to the real and imaginary parts of a non-zero holomorphic 2-form on M via the de Rham isomorphism.

There are two important questions concerning K3 surfaces. One is whether the period map for marked K3 surfaces is surjective, i.e. whether every element of Ω is a period of some marked K3 surface. Another is whether every K3 surface is Kähler. These questions were recently answered in the affirmative by using Yau's proof [69] of the Calabi conjecture (Todorov [62], Looijenga [31], Siu [51, 53]). Prior to Yau's proof of the Calabi conjecture, Hitchin [24] observed that from a Kähler–Einstein metric on a K3 surface one can construct a family of complex structures parametrized by the Riemann sphere. The surjectivity proof depends on this construction of Hitchin. The Kähler property of K3 surfaces follows from a refined version of the surjectivity result, the Torelli theorem for algebraic or Kähler K3 surfaces [8, 32, 42], and the existence of a real closed 2-form with positive $(1, 1)$-component on any K3 surface obtained by Sullivan's trick [60] of using the separation theorem for locally convex linear topological spaces. The simpler surjectivity proofs given in [31, 51] are mainly technical streamlining of the original surjectivity proof of Todorov [62].

The Kähler property of K3 surfaces completes the proof of the conjecture of Kodaira that every compact complex surface with even first Betti number must be Kähler. The other cases were already verified by Miyaoka [34]. An alternative proof of Miyaoka's result was recently given by Harvey and Lawson [23] as an application of their newly obtained intrinsic characterization of Kähler manifolds.

References

[1] Anderson M., The Dirichlet Problem at Infinity for Manifolds of Negative Curvature, *J. Diff. Geom.* **18** (1983), pp. 701–721.

[2] Andreotti A. and Grauert H., Algebraische Körper von automorphen Funktionen, *Nachr. Akad. Wiss. Göttingen, math.-phys. Klasse* (1961), pp. 39–48.

[3] Baily W. L., Jr. and Borel A., Compactification of Arithmetic Quotients of Bounded Symmetric Domains, *Ann. of Math.* **84** (1966), pp. 442–528.

[4] Bando S., On Three-Dimensional Compact Kähler Manifolds of Nonnegative Bisectional Curvature, *J. Diff. Geom.* (1984).

[5] Berger M., Sur les variétés d'Einstein compacts, *C. R. III^e Réunion Math. Expression Latine*, Namur, 1965, pp. 35–55.

[6] Borel A., On the Curvature Tensor of the Hermitian Symmetric Manifolds, *Ann. of Math.* **71** (1960), pp. 508–521.

[7] Burns D., Curvatures of Monge–Ampère Foliations and Parabolic Manifolds, *Ann. of Math.* **115** (1982), pp. 349–373.
[8] Burns D. and Rapoport M., On the Torelli Problem for Kähler K3 Surfaces, *Ann. Scient. Éc. Norm. Sup.* **8** (1975), pp. 235–274.
[9] Burns D., Shnider S., and Wells R. O., On Deformations of Strictly Pseudoconvex Domains, *Invent. Math.* **46** (1978), pp. 237–253.
[10] Calabi E. and Vesentini E., On Compact Locally Symmetric Kähler Manifolds, *Ann. of Math.* **71** (1960), pp. 472–507.
[11] Deligne P. and Mostow G. D., Hypergeometric Functions and Nonarithmetic Monodromy Groups.
[12] Douady A., Le problème des modules pour les sous-espaces analytiques d'un espace analytique donné, *Ann. Inst. Fourier (Grenoble)* **16** (1966), pp. 1–95.
[13] Eells J. and Sampson J. H., Harmonic Mappings of Riemannian Manifolds, *Amer. J. Math.* **86** (1964), pp. 109–160.
[14] Frankel T., Manifolds with Positive Curvature, *Pacific J. Math.* **11** (1961), pp. 165–174.
[15] Gray A., Compact Kähler Manifolds with Nonnegative Sectional Curvature, *Invent. Math.* **41** (1977), pp. 33–43.
[16] Greene R. and Krantz S., Deformation of Complex Structures, Estimates for the $\bar{\partial}$ Equation, and Stability of the Bergman Kernel, *Advances in Math.* **43** (1982), pp. 1–86.
[17] Greene R. and Wu H., Some Function-Theoretic Properties of Noncompact Kähler manifolds, *Proc. Symp. Pure Math.*, Vol. 27, Part 2, Amer. Math. Soc., Providence, R. I. 1975, pp. 33–41.
[18] Greene R. and Wu H., Analysis on Noncompact Kähler Manifolds, *Proc. Symp. Pure Math.*, Vol. 30, Part 2, Amer. Math. Soc., Providence, R. I. 1977, pp. 69–100.
[19] Greene R. and Wu H., Function Theory on Manifolds which Possess a Pole, *Lecture Notes in Math.* **699**, Springer, Berlin–Heidelberg-New York, 1979.
[20] Greene R. and Wu H., Gap Theorems for Noncompact Riemannian Manifolds, *Duke Math. J.* **49** (1982), pp. 731–756.
[21] Gromov M., unpublished.
[22] Hamilton R. S., Three-Manifolds with Positive Ricci Curvature, *J. Diff. Geom.* **17** (1982), pp. 255–306.
[23] Harvey F. R. and Lawson B., An Intrinsic Characterization of Kähler Manifolds, *Invent. Math.* **74** (1983), pp. 169–198.
[24] Hitchin N., Compact Four-Dimensional Einstein Manifolds, *J. Diff. Geom.* **9** (1974), pp. 435–441.
[25] Howard A. and Smyth B., Kähler Surfaces of Nonnegative Curvature, *J. Diff. Geom.* **5** (1971), pp. 491–502.
[26] Howard A., Smyth B., and Wu H., On Compact Kähler Manifolds of Nonnegative Bisectional Curvature I, *Acta Math.* **147** (1981), pp. 51–56.
[27] Kalka M., Deformation of Submanifolds of Strongly Negatively Curved Manifolds, *Math. Ann.* **251** (1980), pp. 243–248.
[28] Kobayashi S. and Ochiai T., Compact Homogeneous Complex Manifolds with Positive Tangent Bundle. In: S. Kobayashi, M. Obata, and T. Takahashi (eds.), *Differential Geometry, in Honor of K. Yano*, Kinokuniya, Tokyo, 1972, pp. 233–242.
[29] Kobayashi S. and Ochiai T., Characterizations of Complex Projective Spaces and Hyperquadrics, *J. Math. Kyoto Univ.* **13** (1973), pp. 31–47.

[30] Kuranishi M., On Locally Complete Families of Complex Analytic Structures, *Ann. of Math.* **75** (1962), pp. 536–577.

[31] Looijenga E., A Torelli Theorem for Kähler–Einstein K3 Surfaces, *Lecture Notes in Math.* **892**, Springer, Berlin–Heidelberg–New York, 1980, pp. 107–112.

[32] Looijenga E. and Peters C., Torelli Theorems for Kähler K3 Surfaces, *Compositio Math.* **42** (1981), pp. 145–186.

[33] Mabuchi T., C^3-Actions and Algebraic Three-Folds with Ample Tangent Bundle, *Nagoya Math. J.* **69** (1978), pp. 33–64.

[34] Miyaoka Y., Kähler Metric on Elliptic Surfaces, *Proc. Japan Acad.* **50** (1974), pp. 533–536.

[35] Mok N., An Embedding Theorem of Complete Kähler Manifolds of Positive Bisectional Curvature onto Affine Algebraic Varieties, *Bull. Soc. Math. France* (1984).

[36] Mok N., Siu Y.-T., and Yau S.-T., The Poincaré–Lelong Equation on Complete Kähler Manifolds, *Compositio Math.* **44** (1981), pp. 183–218.

[37] Mori S., Projective Manifolds with Ample Tangent Bundles, *Ann. of Math.* **110** (1979), pp. 593–606.

[38] Mostow G. D., Strong Rigidity of Locally Symmetric Spaces, *Ann. of Math. Studies* **78** (1973), Princeton University Press.

[39] Mostow G. D., Existence of a Nonarithmetic Lattice in SU(2, 1), *Pacific J. Math.* **86** (1980), pp. 171–276.

[40] Mostow G. D. and Siu Y.-T., A Compact Kähler Surface of Negative Curvature not Covered by the Ball, *Ann. of Math.* **112** (1980), pp. 73–111.

[41] Nakano S., On Complex Analytic Vector Bundles, *J. Math. Soc. Japan* **7** (1955), pp. 1–12.

[42] Piatetskii–Shapiro I. I. and Shafarevitch I. R., A Torelli Theorem for Algebraic Surfaces of Type K3, *Izv. Akad. Nauk SSSR* **35** (1971), pp. 547–572.

[43] Picard E., Sur les fonctions hyperfuchsiennes provenant des séries hypergéometriques de deux variables, *Ann. Ec. Norm.* **2** (1885), pp. 357–384.

[44] Sachs J. and Uhlenbeck K., The Existence of Mininal Immersions of 2-Spheres, *Ann. of Math.* **113** (1981), pp. 1–24.

[45] Satake I., On Compactification of the Quotient Spaces for Arithmetically Defined Discontinuous Groups. *Ann. of Math.* **72** (1960), pp. 555–580.

[46] Schoen R., unpublished.

[47] Siu Y.-T., Some Remarks on the Complex-Analyticity of Harmonic Maps, *Southeast Asian Bull. Math.* **3** (1979), pp. 240–253.

[48] Siu Y.-T., The Complex-Analyticity of Harmonic Maps and the Strong Rigidity of Compact Kähler Manifolds, *Ann. of Math.* **112** (1980), pp. 73–111.

[49] Siu Y.-T., Curvature Characterization of Hyperquadrics, *Duke Math. J.* **47** (1980), pp. 641–654.

[50] Siu Y.-T., Strong Rigidity of Compact Quotients of Exceptional Bounded Symmetric Domains, *Duke Math. J.* **48** (1981), pp. 847–871.

[51] Siu Y.-T., A Simple Proof of the Surjectivity of the Period Map of K3 Surfaces, *Manuscripta Math.* **35** (1981), pp. 311–321. Erratum at the end of [53].

[52] Siu Y.-T., Complex-Analyticity of Harmonic Maps, Vanishing and Lefschetz Theorems, *J. Diff. Geom.* **17** (1982), pp. 55–138.

[53] Siu Y.-T., Every K3 Surface is Kähler, *Invent. Math.* **73** (1983), pp. 139–150.

[54] Siu Y.-T. and Yang, P., Compact Kähler–Einstein Surfaces of Nonpositive Bisectional Curvature, *Invent. Math.* **64** (1981), pp. 471–487.

[55] Siu Y.-T. and Yau S.-T., Complete Kähler Manifolds with Nonpositive Curvature of Faster than Quadratic Decay, *Ann. of Math.* **105** (1977), pp. 255–264. Erratum: *Ann. of Math.* **109** (1979), pp. 621–623.
[56] Siu Y.-T. and Yau S.-T., Compact Kähler Manifolds of Positive Bisectional Curvature, *Invent. Math.* **59** (1980), pp. 189–204.
[57] Siu Y.-T. and Yau S.-T., Compactification of Negatively Curved Complete Kähler Manifolds of Finite Volume, *Ann. of Math. Studies* **102** (1982), pp. 363–380.
[58] Siu Y.-T. and Yau S.-T., in preparation.
[59] Stoll W., The Characterization of Strictly Parabolic Manifolds, *Ann. Scuola Norm. Sup. Pisa* **7** (1980), pp. 87–154.
[60] Sullivan D., Cycles for the Dynamical Study of Foliated Manifolds and Complex Manifolds, *Invent Math.* **36** (1976), pp. 225–255.
[61] Sullivan D., The Dirichlet Problem at Infinity for a Negatively Curved Manifold, *J. Diff. Geom.* **18** (1983), pp. 723–732.
[62] Todorov A. N., Applications of the Kähler–Einstein–Calabi–Yau Metric to Moduli of K3 Surfaces, *Invent. Math.* **61** (1980), pp. 251–265.
[63] Wong B., Characterization of the Unit Ball in C^n by Its Automorphism Group, *Invent. Math.* **41** (1977), pp. 253–257.
[64] Wong B., A Class of Compact Complex Manifolds with Negative Tangent Bundles. In: *Proc. Several Complex Variables Conf. in Madison*, Wisconsin, 1982.
[65] Wong P.-M., Geometry of the Complex Homogeneous Monge–Ampère Equation, *Invent. Math.* **67** (1982), pp. 261–274.
[66] Wu H., Negatively Curved Kähler Manifolds, *Notices Amer. Math. Soc.* **14** (1967), p. 515.
[67] Wu H., On Compact Kähler Manifolds of Nonnegative Bisectional Curvature II, *Acta Math.* **147** (1981), pp. 57–70.
[68] Yang P., On Kähler Manifolds with Negative Holomorphic Bisectional Curvature, *Duke Math. J.* **43** (1976), pp. 871–874.
[69] Yau S.-T., On the Ricci Curvature of a Compact Kähler Manifolds and the Complex Monge–Ampère Equation I, *Comm. Pure Applied Math.* **31** (1978), pp. 339–411.
[70] Zhong J.-C., The Degree of Strong Nondegeneracy of the Bisectional Curvature of Exceptional Bounded Symmetric Domains. In: *Proc. Intern. Conf. Several Complex Variables*, Hangzhou, 1981.

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY,
CAMBRIDGE, MASSACHUSETTS 02138, U.S.A.

Invited 45-Minute Addresses in Sections

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

GREGORY L. CHERLIN

Totally Categorical Structures

§ 1. Introduction

If an axiomatic theory T has a unique model (up to isomorphism) of a given cardinality $\varkappa$, it is said to be *$\varkappa$-categorical*, and if this is the case for all infinite $\varkappa$, then the theory is said to be *totally categorical*. Thus the theory of dense linear orderings having no first or last element is $\aleph_0$-categorical, the theory of algebraically closed fields of specified characteristic is $\aleph_1$-categorical, and a slightly non-trivial example of a totally categorical theory is provided by the theory $A(p^2)$ of abelian groups of exponent p^2 in which every element of order p is divisible by p.

These three examples are known to be typical in many respects, though the full story is by no means known. As a simple example, it is well known that for any countable model Q of an $\aleph_0$-categorical theory, the automorphism group G of Q has only finitely many orbits on Q^n (for any n). This in fact characterizes $\aleph_0$-categorical theories; but the analysis of $\varkappa$-categorical theories for $\varkappa$ uncountable is more elaborate (as the natural example suggests).

Good results on totally categorical theories in general have only become available quite recently, and the picture is still quite incomplete. The main results are:

THEOREM I. *A totally categorical structure*[1] *A has the finite submodel property: any first order property of A is also a property of some finite substructure of A.*

THEOREM II. *Any totally categorical structure A can be coordinatized by a collection of geometries (affine, projective, or degenerate).*

---

[1] By a convenient abuse of terminology, this means, "a model of a totally categorical theory".

A precise version of this is in § 2. When A is $(\mathbf{Z}/p^2\mathbf{Z})^{(\infty)}$, and $O(p^i)$ $(i=0,1,2)$ are the orbits in A under $\operatorname{Aut} A$ (elements of order p^i), then $O(p)$ becomes a projective geometry after factoring out an equivalence relation with finite classes, and $O(p^2)$ is fibered over $O(p)$ by affine geometries. The example is typical.

The proofs outlined in § 2 depend indirectly on the finiteness of the set of sporadic finite simple groups. This can be avoided (§ 3).

In the remainder of the paper, I use customary model-theoretical terminology.

§ 2. The structure theory

This outline follows [2], with terminological deviations. All structures will be $\aleph_0$-categorical and $\aleph_0$-stable.

Terminology. An infinite *Jordan geometry* is a set S equipped with a permutation group G such that for every finite subset $X \subseteq S$:

(J) The pointwise stabilizer $G(X)$ of X in G has finitely many orbits, exactly one of which is infinite.

Example. If S is a strongly minimal $\aleph_0$-categorical structure and $G = \operatorname{Aut} S$, then (S, G) is an infinite Jordan geometry.

Jordan geometries are indeed geometries: one can define the closure $\langle X\rangle$ of a finite set X as the union of the finite orbits of $G(X)$, and take the closed sets as subspaces. The geometry is *primitive* if points are closed, and any geometry can be made primitive by removing $\langle \emptyset\rangle$ and passing to a quotient.

PROPOSITION 1. *Infinite primitive Jordan geometries are affine, projective, or degenerate.*

The proof is easy, since such a geometry is a limit of finite geometries with a doubly transitive automorphism group, and all large doubly transitive finite groups are known as a consequence of the finiteness of the set of sporadic finite simple groups. (For references see [1] or [2].)

A strongly minimal set associated with a projective or degenerate primitive geometry will be called *modular*. Two strongly minimal sets contained in a given structure M are *orthogonal* if the introduction of constants naming the elements of one set has no effect on the geometry of the other (as substructures of M). The following result is very powerful, since orthogonal strongly minimal sets are easy to handle.

PROPOSITION 2. *There is a 0-definable bijection between the elements of any two nonorthogonal strongly minimal $\aleph_0$ categorical primitive modular structures.*

Formulated as a slogan, this becomes: All difficulties are caused by affine geometries.

More terminology (cf. the example following Proposition 3):

1. The structure M is transitive if Aut M is.

2. If M, $A \subseteq M^*$ with M and A transitive, then M is *coordinatized* by A (in M^*) if:

$$\forall x \in M \quad \langle x \rangle \cap A \neq \emptyset. \tag{$*$}$$

3. A is attached to M if there is a structure M^* 0-interpretable in M such that $M \cup A \subseteq M^*$. (The underlying set of M^* is a 0-definable subset of M^n for some n, taken modulo a 0-definable equivalence relation.)

PROPOSITION 3. *Any $\aleph_0$-categorical, $\aleph_0$-stable, transitive M of finite rank n can be coordinatized by a rank one set attached to it.*

(*Example.* Let M be the set of all unordered pairs of distinct elements of a set S, equipped with the binary relation R defined by: $|p_1 \cap p_2| = 1$. Then $R \subseteq M^2$, and the equivalence relation E on R defined by: $\big((p_1, p_2), (p_3, p_4)\big) \in E$ iff $p_1 \cap p_2 = p_3 \cap p_4$ is easily seen to be definable from R. R/E is a strongly minimal set attached to and coordinatizing M; it can be identified with S.)

This is a key result. For the proof, pick a rank $(n-1)$ set S whose definition involves parameters $\bar{a}$. We may assume S is normalized in the sense of [3], and varying the parameters get a transitive family A of such sets, with $\bigcup A = M$. The main point is then that rank $A = 1$ (which is perhaps unexpected) and it then follows easily that A coordinatizes M. (It is routine to attach A to M.)

Using this result, one proves:

PROPOSITION 4. *If M is $\aleph_0$-categorical and $\aleph_0$-stable, then M has finite rank.*

Theorem II follows.

Theorem I can be derived using Theorem II. If $D \subseteq M$ is a strongly minimal set and $X \subseteq D$ is nonempty, an *envelope* $E(X)$ is a maximal subset of M independent from D over X.

PROPOSITION 5. *If $M \models \psi$ and X is large enough, $E(X) \models \psi$.*

PROPOSITION 6. *If X is finite then rank $E(X) <$ rank M. By repeated application of these results, we arrive at Theorem I, generalized to $\aleph_0$-categorical $\aleph_0$-stable structures.*

The proof of Proposition 5 consists of an appropriately modified Tarski–Vaught test. If $M \models \psi(a)$ where the (few) parameters of ψ all lie in $E(X)$, one can proceed by induction on rank $(a/E(X))$, to find a similar element in $E(X)$. Theorem II is used here.

§ 3. Historical remarks

It was conjectured in ancient times that a complete $\aleph_1$-categorical theory is never finitely axiomatizable. The conjecture decomposes as follows:

(A) A totally categorical theory has the finite submodel property.

(B) An $\aleph_1$-categorical theory which is not $\aleph_0$-categorical is never finitely axiomatizable.

Peretyat'kin refuted (B) in [10]; subsequently both Morley and Parigot came up with noticeably simpler variants (unpublished). The conjecture (A) was proved by Zil'ber in [12] modulo a serious gap caused by over-simplifying the application of Lemma 5.1. The gap can be filled by a number-theoretic argument. (As far as I know, Zil'ber has not yet published this correction.)[2] Much of the material in [13] was already in [12], in less elegant form (the Parisian viewpoint [6, 7] was taking root in the meantime).

The work in [2] began when the gap in [13] still seemed serious. Lachlan had pointed out the relevance of certain incidence geometries called *pseudoplanes* to some conjectures about $\aleph_0$-categorical theories [3], and Zil'ber established their relevance to the classification of strongly minimal $\aleph_0$-categorical structures in [16 I]. The model-theoretic form of Proposition 1 was conjectured in [16 I], and proved in [16 II, III] by a direct method using a lot of model theory. This last work dates to summer 1980.

I studied [13] in 1980 during Model Theory Year at the I.A.S. in Jerusalem. In the fall I noticed the truth and relevance of Proposition 1, having just heard of the completion of the classification of finite simple groups. As I was very conveniently surrounded by model theorists at the time, it didn't take very long for the picture presented here to come into focus. (A graduate student in logic at Wisconsin noticed the truth of Proposition 1 about the same time [9].)

My exposition slights the detailed theory of envelopes [2, § 7], which has been pushed further in [8]. The notion was introduced in [13]. Lachlan came upon a particularly transparent special case in his study of stable

[2] See Note 2.

structures admitting elimination of quantifiers in a finite relational language [4], a special case of "disintegrated" structures ($\aleph_0$-categorical, $\aleph_0$-stable, with all coordinatizing geometries degenerate). In particular, the fundamental conjecture that $\aleph_0$-categorical $\aleph_0$-stable structures should be infinite analogs of certain finite structure can be made entirely precise in this context, giving the main conjecture of [4], proved for binary languages in [5].

Continuing this historical outline into the near future, a proof of this conjecture based on finite permutation group theory of the type outlined in [1] is under construction (January 1983). If successful, it should provide further evidence of the usefulness of group-theoretic methods in this area of model theory.

The geometric analysis of sets of rank a inside a set of rank $a+1$, which dominates the proof of Proposition 3, is being pursued by Buechler.

Finally, the classification of the primitive geometries associated with strongly minimal $\aleph_0$-categorical structures yields a classification of the structures themselves if singletons are algebraically closed (the primitive case). In the imprimitive case, the Galois group of the underlying field may act on the blocks of imprimitivity causing complications which are not understood, but is it conjectured that strongly minimal $\aleph_0$-categorical structures in which the Galois group does not act (i.e. field elements are named) should be classifiable. (This is true when the quotient is degenerate.)

References

[1] Cameron P., Finite Permutation Groups and Finite Simple Groups, *BLMS*, **13** (1981), pp. 1–22.

[2] Cherlin G., Harrington L. and Lachlan A., $\aleph_0$-Categorical, $\aleph_0$-Stable Structures, to appear in *Ann. Pure Appl. Logic.*

[3] Lachlan A., Two Conjectures Regarding the Stability of ω-Categorical Theories, *Fund. Math.*, **81** (1974), pp. 133–145.

[4] Lachlan A., *On Countable Stable Structures Which are Homogeneous for a Finite Relational Language*, to appear.

[5] Lachlan A. and Shelah S., *Stable Structures Homogeneous for a Finite Binary Language*, to appear.

[6] Lascar D., Ranks and Definability for Super Stable Theories, *Israel J. Math.*, **23** (1976), pp. 53–87.

[7] Lascar D. and Poizat B., An Introduction to Forking, *JSL*, **44** (1979), pp. 330–350.

[8] Loveys J., *The Uniqueness of Envelopes in $\aleph_0$-Categorical, $\aleph_0$-Stable Structures*, M.Sc. thesis, Simon Fraser University, 1982.

[9] Mills C., *$\aleph_0$-Categorical Strongly Minimal Sets*, talk at the Midwest Model Theory Conference, Urbana, Spring, 1981.

[10] Peretyat'kin M., An Example of an ω_1-Categorical Complete Finitely Axiomatizable Theory, English translation: *Alg. and Logic*, **19** (1980), pp. 202–229.
[11] Shelah S., *Classification Theory*, North-Holland, Amsterdam, 1978.
[12] Zil'ber B., *The Structure of Models of Categorical Theories and the Problem of Axiomatizability* (in Russian), VINITI, Dep. No. 2800–77 (1977), mimeo, ca. 80 pp.
[13] Zil'ber B., Totally Categorical Theories: Structural Properties and the Nonfinite Axiomatizability. In: Pacholski *et al.* (eds.), *Model Theory of Algebra and Arithmetic*, LNM 834, Springer-Verlag, NY, 1980, pp. 381–410.
[14] Zil'ber B., On the Problem of Finite Axiomatizability for Theories Categorical in all Finite Powers. In: B. Baizanov (ed.), *The Theory of Models and its Applications* (in Russian), Kazah. Gos. Univ., Alma-Ata, 1980, MR 82g: 03051, pp. 47–60.
[15] Zil'ber B., *Totally Categorical Structures and Combinatorial Geometries* (in Russian), English translation: *Soviet Math. Doklady*, **259** (1981), pp. 149–151.
[16] Zil'ber B., Strongly Minimal Countably Categorical Structures I–III (in Russian), *Sibirski Mat. Zh.*, I, **21** (1980), pp. 98–112; II–III, to appear.

Notes

1. Zoe Chatzidakis has shown that profinite groups with the Iwasawa ("embedding") property are naturally associated with $\aleph_0$-categorical $\aleph_0$-stable many-sorted structures, to which a slight generalization of the theory described here applies.

2. The correction to [13] referred to above has been published in:

[17] Zil'ber B., On the Problem of Finite Axiomatizability for Theories Categorical in all Infinite Powers. In: B. Baižanov (ed.), *Researches in Theoretical Programming*, Kazah. Gos. Univ., Alma-Ata, 1981, pp. 69–75.

3. The conjecture in [4] has been proved in the manner anticipated.

RUTGERS UNIVERSITY

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

JEAN-YVES GIRARD

The Ω-Rule

1. Background: the β-rule

1.1. Among the contributions of Poland to mathematical logic, one of the most original and far-reaching is the theory of so-called *β-models*, introduced by Mostowski in the early 60's: A β-model of the second order arithmetic is a model where well-foundedness is absolute, i.e.

$$M \vDash WF[\bar{C}] \leftrightarrow \{(n, m);\ M \vDash \bar{n}\bar{C}\bar{m} \text{ is well-founded}\}.$$

This notion (together with the concept of ω-model) is one of the genuine improvements of the concept of model, closely related to the concept of transitive model in set theory. A famous result of Mostowski and Suzuki states that we cannot by any means reduce β-models to ω-models.

1.2. In analogy to the well-known completeness theorem of Orey for ω-logic, Mostowski raised the question of characterizing validity in all β-models by a system of axioms and rules: the *β-rule.* The answer is by no means trivial: truth in all models is Σ^0_1-complete, truth in all ω-models is Π^1_1-complete and truth in all β-models is Π^1_2-complete; hence we must find, if we want to solve Mostowski's problem, a notion of proof drastically different from the finitary trees and the well-founded ω-branching trees which answer the completeness problem for the usual models and the ω-models respectively. In the middle 70's, Apt was able to show that no solution is possible without introducing a completely new idea.

1.3. A more general notion of β-model is the following: if $\mathscr{T}$ is a denumerable theory in a language L containing a distinguished type 0 for ordinals, and a binary relation $<$ on that type, a β-model of $\mathscr{T}$ is a model of $\mathscr{T}$ such that $|M| = (M(0), M(<))$ is an ordinal. ($M(0) = \emptyset$ is allowed.) Mostowski's notion is easily reduced to the new one, so we shall now only deal with that one. Assume that the formula A is true in all the β-models of $\mathscr{T}$ such that $|M| = x$. It follows immediately (at least when x is denumerable) that A has an x-proof in $\mathscr{T}$, i.e. a proof making use of the *x-rule*:

$$\frac{\Gamma \mid\!\rightarrow A[\bar{o}], \Delta \ldots \Gamma \mid\!\rightarrow A[\bar{z}], \Delta \ldots \text{all } z < x}{\Gamma \mid\!\rightarrow \forall a^0 A[a], \Delta}$$

So, the truth of A in all β-models of $\mathscr{T}$ implies the existence of a family $(\pi_x)_{x \in On}$ such that for all ordinals x, π_x is an x-proof of A in $\mathscr{T}$. Such a family can be considered as a β-proof of A in $\mathscr{T}$. However, as it stands, our solution is both trivial and ridiculous: a proof must be a "syntactic" object, i.e., it must be finitarily graspable in some way! A family of infinitary proofs, indexed by the class of all ordinals (or even by denumerable ones) is by no means graspable. But the solution would not be ridiculous if we were able to generate our family (π_x) from a reasonable denumerable set of data, and to do this in an effective way!

1.4. If $x, x' \in On$ and $f \in \mathrm{I}(x, x')$ (i.e., f is a strictly increasing function from x to x'), if π is an x'-proof, then one can (in certain cases) define an inverse image $f^{-1}(\pi)$ of π under f; $f^{-1}(\pi)$ will be an x-proof. This process is called *mutilation* and runs in two steps:

(i) In π delete all sequents occurring above any premise of any index $z \notin \mathrm{rg}(f)$ of any x'-rule.

(ii) If in the remaining "proof" all ordinals which occur are in $\mathrm{rg}(f)$, then replace systematically parameters $\overline{f(z)}$ by $\bar{z}$.

A *β-proof* of A in $\mathscr{T}$ is a family $(\pi_x)_{x \in On}$ such that:

(i) For all x, π_x is an x-proof of A in $\mathscr{T}$,

(ii) For all x, x' and $f \in \mathrm{I}(x, x') f^{-1}(\pi_{x'}) = \pi_x$.

An alternative formulation is to present a β-proof as a functor from the category ON of ordinals into a category $\mathrm{DEM}_{\mathscr{T}}$ of ordinal-branching proofs. Then one easily checks that such a functor preserves direct limits and pull-backs; and, since any ordinal is a direct limit of integers, the subfamily $(\pi_n)_{n<\omega}$ determines the complete family $(\pi_x)_{x<On}$[1] and does this

[1] But given any family $(\pi_n)_{n \in On}$ such that $f \in \mathrm{I}(n, m) \rightarrow f^{-1}(\pi_m) = \pi_n$, the extensions $(\pi_x)_{x \in On}$ need not be well-founded for all $x \in On$; similarly an analytic function defined on a neighbourhood of 0 may have no extension to the disc $|z| < 1$ (but the extension, if it exists, is uniquely and effectively determined).

in a completely effective way. Since the proofs π_n are obviously finite, it makes sense to speak of a recursive family $(\pi_n)_{n \in On}$: we can therefore define the concept of a *recursive* β-proof, and we have obtained a reasonable syntactic notion.

1.5. I introduced the concepts of 1.4 in 1978 and proved *completeness*: A is true in all β-models of $\mathscr{T}$ iff there is a β-proof of A in $\mathscr{T}$. Furthermore, when $\mathscr{T}$ is recursive, the β-proof can be chosen recursive. The result gave a complete solution to Mostowski's problem, and had a lot of applications in generalized recursion. The reason for these applications is quite simple: in many situations of admissible set-theory, the structure we are dealing with is the only β-model of a finite theory $\mathscr{T}$; combining completeness with cut-elimination techniques, I was able to reduce generalized recursion over reasonable successor-admissible ordinals to usual recursion. These results were simplified and/or improved by Masseron, Ressayre, Van de Wiele, Normann, Vauzeilles, Jäger, Buchholz,...

1.6. Because of its interesting semantics and its nice syntax, β-logic is the most natural generalization of ω-logic. We are dealing with large objects (the ordinals x in π_x are arbitrarily large), but we keep a finitary control (by means of the π_n's). Among the important formal properties of β-logic, let us mention:

(i) *Interpolation*: an interpolation theorem was obtained by Vauzeilles; similarly to ω-logic, we must enlarge the language: it is necessary to introduce $L_{\beta\omega}$ where the formulas themselves are functors; in other terms, we allow conjunctions of variable lengths.

(ii) *Cut-elimination*: $L_{\beta\omega}$ enjoys a reasonable form of cut-elimination. Ferbus obtained bounds for the cut-elimination process by means of the functorial version of the Veblen hierarchy.

(iii) The underlying structure behind β-proofs, when endowed with the familiar Kleene–Brouwer ordering, leads to the concept of a *dilator*, fundamental in *Π^1_2-logic*, i.e., in the theory of all notions related to β-proofs and their applications. There is a close connection between syntactic operations on β-proofs and the natural properties of dilators.

1.7. However, β-logic, as it stands, is not completely satisfactory; in particular, we would like to make a β-proof appear as a (well-founded) succession of rules, the usual ones plus a specific one, the β-rule. The need for a specific β-rule arises from the fact that the most advanced techniques on the subject consist in cut-elimination theorems for theories of inductive definitions, in which a presentation of the deductive framework by means of axioms and rules is simply crucial. It is possible to find such a presentation for β-logic (see [1] and also the unpublished Ch. 2 of Vauzeilles *Thèse d'Etat*); but the solution thus found is not completely satisfactory.

1.8. *Ω-logic* is the last word (?) on Mostowski's problem: by replacing the category ON by the wider WF, we obtain a greater flexibility (for instance WF has direct sums), and in particular a very satisfactory specific Ω-rule. (We have chosen "Ω" to stress both the analogy with and the difference from the case of ω-logic.[2]) The theory of the underlying structures behind Ω-proofs is very rich: these objects are called *gerbes* and seem to be an adequate framework for a general geometry of ordinals.

2. The languages $L_{R,\vec{a}}$

2.1. In the rest of the paper, L is a fixed first order language; we shall save space by assuming that the formulas of L are built from atomic ones by means of the unique operator $QxAB$ (whose intended meaning is $\forall x(\sim A \ \& \sim B)$. The usual connectives and quantifiers are easily defined from Q.

2.2. WF is the category of *well-founded orders*:

$R \in |\mathrm{WF}|$ iff R is a binary relation (defined on a set $|R|$) which is irreflexive, transitive, and well-founded: there is no sequence $(a_n)_{n\in N}$ in $|R|$ such that $a_{n+1}Ra_n$ for all $n \in N$.

$\varrho \in \mathrm{WF}(R, S)$ iff ϱ is a function from $|R|$ to $|S|$ such that $aRb \to \varrho(a)S\varrho(b)$. FIN is the full subcategory of WF consisting of finite orders.

2.3. A *direct system*[3] $(R_\iota, \varrho_{\iota\varkappa})$ admits (R, ϱ_ι) as its direct limit in WF iff:

(i) $\varrho_\iota \in \mathrm{WF}(R_\iota, R)$,

(ii) $\iota < \varkappa \to \varrho_\iota = \varrho_\varkappa \varrho_{\iota\varkappa}$,

(iii) If $a \in |R|$, then $a \in \mathrm{rg}(\varrho_\iota)$ for some index ι,

(iv) If aRb there are some index ι and some $a', b' \in |R_\iota|$ s.t. $a = \varrho_\iota(a')$, $b = \varrho_\iota(b')$ and $a'R_\iota b'$.

From this we see that every object of WF is the direct limit of a direct system in FIN. But most of direct systems in FIN have no direct limit in WF. In order to get a category containing WF and closed under direct

[2] The "Ω-rule" considered here has no known relation to the rule of the same name considered by Buchholz.

[3] A "direct system" means an inductive system indexed by a nonvoid directed ordered set.

limits, one must introduce the category **OR** of strict orders, i.e., drop the well-foundedness condition in 2.2.

2.4. (i) If $R \in |\mathrm{WF}|$, we define $R+1$ by $|R+1| = |R| \cup \{\bar{R}\}$; $aR+1\,b$ iff aRb when $a, b \in |R|$; $aR+1\,\bar{R}$ when $a \in |R|$. When $\varrho \in \mathrm{WF}(R, S)$, we define $\varrho+1 \in \mathrm{WF}(R+1, S+1)$ by $\varrho+1(a) = \varrho(a)$ for $a \in |R|$, $\varrho+1(\bar{R}) = \bar{S}$.

(ii) If $R \in |\mathrm{WF}|$ and $i \in |R|$, we define R^*i by $|R^*i| = |R| \cup \{j\}$ (with $j \notin |R|$) aR^*ib iff aRb when $a \in |R|$, $b \in |R|$, jR^*ia iff $i = a$ or iRa. If $\varrho \in \mathrm{WF}(R, S)$, if $i \in |R|$ and $bS\varrho(i)$, we define $\varrho^*ib \in \mathrm{WF}(R^*i, S)$ by:

$$\varrho^*ib(a) = \varrho(a) \quad \text{for } a \in |R|, \quad \varrho^*ib(j) = b.$$

2.5. If $K \in |\mathrm{WF}|$, we define the category WF^K by:
$|\mathrm{WF}^K| = \{(R, d);\ d \in \mathrm{WF}(K, R)\}$,
$\mathrm{WF}^K(R, d; S, e) = \{\varrho \in \mathrm{WF}(R, S);\ e = \varrho d\}$.

2.6. Assume that $(R, d) \in |\mathrm{WF}^K|$; we then define the language $L_{R,d}$ inductively:

(i) If A is an atomic formula of L, then $A \in L_{R,d}$,

(ii) If $A, B \in L_{R,d}$ and x is a variable of L, then $QxAB \in L_{R,d}$,

(iii) If $i \in |K|$, assume that $(A_b)_{bRd(i)}$ is a family of formulas, $A_b \in L_{R,d^*ib}$, such that only finitely many variables are free in the whole family; then $\bigwedge\!\!\!\bigwedge_{bRd(i)} A_b \in L_{R,d}$,

(iv) The only formulas of the languages $L_{R,d}$ are those given by (i)–(iii).

2.7. Assume that $\varrho \in \mathrm{WF}^K(R, d; R', d')$ and that $A' \in L_{R',d'}$; then we define $\varrho^{-1}(A') \in L_{R,d}$ by *mutilation*:

(i) $\varrho^{-1}(A') = A'$ if A' is atomic,

(ii) $\varrho^{-1}(QxA'B') = Qx\varrho^{-1}(A')\varrho^{-1}(B')$,

(iii) $\varrho^{-1}(\bigwedge\!\!\!\bigwedge_{b'R'd'(i)} A'_{b'}) = \bigwedge\!\!\!\bigwedge_{bRd(i)} \varrho^{-1}(A'_{\varrho(b)})$: here we remark that

$$\varrho \in \mathrm{WF}^{K^*i}\big(R, d^*ib; R', d'^*i\varrho(b)\big).$$

2.8. *Example.* Assume that $F[x, y]$ is a formula of L; we introduce, for $R \in |\mathrm{WF}|$ and $a \in |R|$, the formulas:

$$Acc(R, a, F, x) =_{\mathrm{def}} \mathop{\mathbf{W}}_{bRa} \forall y\big(F[y, x] \to Acc(R, b, F, y)\big),$$

$$Acc(R, F) =_{\mathrm{def}} \mathop{\mathbf{W}}_{bR+1\bar{R}} \forall x Acc(R, b, F, x).$$

When F is an order and R is a well-order, then $Acc(F, R)$ means that F is a well-order and $\|F\| < \|R\|$. If $|K| = \{o\}$ and $d_a(o) = a$ then $Acc(R, a, F, x) \in L_{R,d_a}$. If $\varrho \in \mathrm{WF}(R, R')$, then

$$\varrho^{-1}\big(Acc(R', \varrho(a), F, x)\big) = Acc(R, a, F, x).$$

$Acc(R, F) \in L_{R+1,\emptyset}$ and, if $\varrho \in \mathrm{WF}(R, R')$ then

$$\varrho + 1^{-1}\big(Acc(R', F)\big) = Acc(R, F).$$

2.9. If $K \in |\mathrm{WF}|$, the category FOR^K is defined by:

$$|\mathrm{FOR}^K| = \{(R, d, A);\ (R, d) \in |\mathrm{WF}^K| \text{ and } A \in L_{R,d}\},$$

$\varrho \in \mathrm{FOR}^K(R, d, A; R', d', A')$ iff $\varrho \in \mathrm{WF}^K(R, d; R', d')$ and $\varrho^{-1}(A') = A$.

2.10. The functor *type*: $\mathrm{t}(R, d, A) = R$, $\mathrm{t}(\varrho) = \varrho$ preserves and reflects direct limits: this means that

$$(R, d, A; \varrho_\iota) = \varinjlim(R_\iota, d_\iota, A_\iota; \varrho_{\iota\varkappa}) \text{ iff } (R, \varrho_\iota) = \varinjlim(R_\iota, \varrho_{\iota\varkappa}).$$

But this functor does not create direct limits: the existence of $\varinjlim(R_\iota, \varrho_{\iota\varkappa})$ in WF is not enough to ensure the existence of $\varinjlim(R_\iota, d_\iota, A_\iota; \varrho_{\iota\varkappa})$ in FOR^K. However, the direct limit will exist in a wider category $\overline{\mathrm{FOR}^K}$, defined as FOR^K, $L_{R,d}$ being replaced by $\bar{L}_{R,d}$: the formulas of $L_{R,d}$ can be viewed as trees; these trees are characterized by local conditions (on the branchings) and a global one, well-foundedness. $\bar{L}_{R,d}$ is obtained by keeping the local conditions and dropping well-foundedness; the elements of $\bar{L}_{R,d}$ are called *preformulas.*

2.11. The category REP is defined by: $|\mathrm{REP}| = |\mathrm{WF}|$, $\varrho \in \mathrm{REP}(R, S)$ iff ϱ is *replete,* i.e., ϱ is surjective and, for all $a \in |R|$, ϱ maps $\{b; bRa\}$ onto $\{b'; b'S\varrho(a)\}$.

2.12. Let us denote by θ_A the tree associated with a formula $A \in L_{R,d}$; if $A = \varrho^{-1}(A')$, then one easily builds a function $\theta_\varrho \in \mathrm{WF}(\theta_A, \theta_{A'})$; this function is injective when ϱ is injective, and is surjective when ϱ is replete.

2.13. Assume that M is a structure for L. Then we easily define the notion of validity of a closed formula $A \in L_{R,d}$ in M, denoted by $M \models A$. One easily checks that, if $\varrho \in \mathrm{REP}(R, R')$ and $A = \varrho^{-1}(A')$, then $M \models A$ iff $M \models A'$. This illustrates the extreme importance of replete morphisms.

2.14. Given $R \in |\mathrm{WF}|$, there are two main replete morphisms connected with R:

(i) the canonical morphism from R to $\|R\|$,

(ii) if T_R denotes the tree of nonvoid descending sequences of R, ordered by the extension relation, the function $\varrho \in \mathrm{REP}(T_R, R)$:

$$\varrho\big((x_o, \ldots, x_n)\big) = x_n.$$

In particular, in the functional languages $L_{\Omega,S,d}$, the meaning of $\Phi(R)$ will only depend an $\|R\|$, so $\Phi(\cdot)$ only speaks about ordinals, and there is no real gain of expressive power w.r.t. β-logic; on the other hand, when x is an ordinal, the formula $\Phi(T_x)$ has nicer geometrical properties than $\Phi(x)$...

3. The calculi $Lk_{R,S,d}$

3.1. In the sequel we shall work with the following stock of axioms and rules:

I. Axioms.

$$A \mid\!\rightarrow A \qquad (A \text{ atomic}).$$

II. Logical rules.

Q-introduction

$$\frac{\Gamma, A \mid\!\rightarrow \Delta \quad \Lambda, B \mid\!\rightarrow \Pi}{\Gamma, \Lambda \mid\!\rightarrow QxAB, \Delta, \Pi}\, rQ^{4}$$

Q-elimination

$$\frac{\Gamma \mid\!\rightarrow A[t/x], \Delta}{\Gamma, QxAB \mid\!\rightarrow \Delta}\, l1Q \qquad \frac{\Gamma \mid\!\rightarrow B[t/x], \Delta}{\Gamma, QxAB \mid\!\rightarrow \Delta}\, l2Q$$

$\bigwedge$-introduction

$$\frac{\Gamma \mid\!\rightarrow A_b, \Delta}{\Gamma \mid\!\rightarrow \bigwedge_{bRa} A_b, \Delta}\, r\!\bigwedge\! Ra$$

$\bigwedge$-elimination

$$\frac{\Gamma, A_{b_o} \mid\!\rightarrow \Delta}{\Gamma, \bigwedge_{bRa} A_b \mid\!\rightarrow \Delta}\, l\!\bigwedge\! Rb_o a$$

In $(l\!\bigwedge\! Rb_o a)$, $b_o Ra$.

III. Structural rules. These rules are the usual ones: weakening (rW) and (lW), exchange (rE) and (lE), and contraction (rC) and (lC).

IV. Cut rule.

$$\frac{\Gamma \mid\!\rightarrow A, \Delta \qquad \Lambda, A \mid\!\rightarrow \Pi}{\Gamma, \Lambda \mid\!\rightarrow \Delta, \Pi}\, CUT$$

3.2. Assume that $R \in |\mathrm{WF}|$ and that $d \in \mathrm{WF}\big(K, R+1\big)$; then we define the sequent calculus $Lk_{R,d}$: the sequents are expressions $\Gamma \mid\!\rightarrow \Delta$, where

[4] x is not free in Γ, $\Lambda \mid\!\rightarrow \Delta$, Π.

Γ and Δ are finite sequences of formulas in $L_{R+1,d}$. The proofs are defined inductively by iterating the rules and axioms given in 3.1; the rules for $\bigwedge\!\!\!\bigwedge$ deserve special mention:

(i) If for all b such that $bRd(i)$, π_b is a proof of $\Gamma \mid\!\!\to A_b, \Delta$ in $Lk_{R,d*ib}$, then we can form a new proof π of $\Gamma \mid\!\!\to \bigwedge\!\!\!\bigwedge_{bRd(i)} A_b, \Delta$ in $Lk_{R,d}$.

(ii) Assume that jKi; then from a proof λ of $\Gamma, A_{d(j)} \mid\!\!\to \Delta$ in $Lk_{R,d}$ we can construct a proof π of $\Gamma, \bigwedge\!\!\!\bigwedge_{bRd(i)} A_b \mid\!\!\to \Delta$ in $Lk_{R,d}$.

(iii) The remaining clauses are all of the following form: assume that $\pi_1, \ldots, \pi_k$ are proofs of $\Gamma_1 \mid\!\!\to \Delta_1, \ldots, \Gamma_k \mid\!\!\to \Delta_k$ in $Lk_{R,d}$ and that the rule or axiom (θ) applies to these sequents, and yields $\Gamma \mid\!\!\to \Delta$; then we can form a new proof π of the sequent $\Gamma \mid\!\!\to \Delta$ in $Lk_{R,S,d}$.

3.3. Assume $\varrho \in \mathrm{WF}(R, R')$ and that $d' = \varrho + 1d$; then, with each proof π' in $Lk_{R',d'}$ we associate a proof

$$\pi = (\varrho, \sigma)^{-1}(\pi')$$

in $Lk_{R,d}$, on the model of 2.7, using the function $\varrho+1$. If the conclusion of π' is $\Gamma' \mid\!\!\to \Delta'$ $(= A'_1, \ldots, A'_n \mid\!\!\to A_{n+1}, \ldots, A'_m)$, then the conclusion of π is $\varrho^{-1}(\Gamma' \mid\!\!\to \Delta')$ $(= A_1, \ldots, A_n \mid\!\!\to A_{n+1}, \ldots, A_m$, with $A_i = \varrho+1^{-1}(A'_i))$.

3.4. We can form a category DEM^K of proofs. Its objects are 3-tuples (R, d, π), with $\pi \in Lk_{R,d}$, and $\varrho \in \mathrm{DEM}^K(R, d, \pi; R', d', \pi')$ iff $\varrho+1 \in \mathrm{WF}^K(R+1, R+1)$ and $\pi = \varrho^{-1}(\pi')$.

3.5. The functor "type": $\mathrm{t}(R, d, \pi) = R$, $\mathrm{t}(\varrho) = \varrho$, preserves and reflects direct limits; but it does not create them. It is easy to see that the analogues of Remarks 2.10 and 2.12 hold for DEM^K.

4. The calculi $L_{\Omega,S,d}$

4.1. Assume that $R, S \in |\mathrm{WF}|$; then we define $R \oplus S$ by:

$$|R \oplus S| = \{o\} \times |R| \cup \{1\} \times |S|,$$

and

$$(i, a) R \oplus S (j, b) \text{ iff } i = j = o \text{ and } aRb \text{ or } i = j = 1 \text{ and } aSb.$$

If $\varrho \in \mathrm{WF}(R, R')$, and if $\sigma \in \mathrm{WF}(S, S')$, we define

$$\varrho \oplus \sigma \in \mathrm{WF}(R \oplus S, R' \oplus S')$$

by

$$\varrho\oplus\sigma(o, a) = (o, \varrho(a)), \quad \varrho\oplus\sigma(1, a) = (1, \varrho(a')).$$

Obviously, $\oplus$ is a direct sum in the category WF.

We make the following notational conventions:

(i) We consider that $\oplus$ is associative and that the void order $\emptyset$ is neutral for $\oplus$.

(ii) If $d \in \mathrm{WF}(K, S)$ and e is the only element of $\mathrm{WF}(\emptyset, R)$, then $(e\oplus d)+1 \in \mathrm{WF}((\emptyset\oplus K)+1, (R\oplus S)+1)$. By (i) we have $\emptyset\oplus K = K$. We shall use the notation $d+1$ instead of $(e\oplus d)+1$, hence

$$d+1 \in \mathrm{WF}(K+1, (R\oplus S)+1).$$

4.2. Assume that $S \in |\mathrm{WF}|$ and $d \in \mathrm{WF}(K, S)$; then we define the language $L_{\Omega,S,d}$ as follows: the formulas are functors Φ from WF to FOR^{K+1} such that:

(i) $\Phi(R) = ((R\oplus S)+1, d+1, \Phi_R)$,

(ii) $\Phi(\varrho) = (\varrho\oplus \mathrm{id}_S)+1$,

(iii) If λ_R is the canonical surjection from $(R\oplus S\oplus S)+1$ to $(R\oplus S)+1$, then $\lambda_R^{-1}(\Phi_R) = \Phi_{R\oplus S}$. (Observe that λ_R is replete.)

4.3. The formulas of $L_{\Omega,S,d}$ preserve direct limits (they also preserve pull-backs and kernels). In particular, such functors are completely and effectively determined by the values $\Phi(R)$ for $R \in |\mathrm{FIN}|$. When S is itself finite, then these formulas are finite.

4.4. The calculus $Lk_{\Omega,S,d}$ is defined as follows: the proof sare functors Π from WF to DEM^{K+1} such that:

(i) $\Pi(R) = (R\oplus S)+1, d+1, \Pi_R)$,

(ii) $\Pi(\varrho) = (\varrho\oplus \mathrm{id}_S)+1$,

(iii) $\lambda_R^{-1}(\Pi(R)) = \Pi(R\oplus S)$

4.5. For proofs in $Lk_{\Omega,S,d}$, remarks analogous to 4.3 are obviously true.

4.6. If the conclusion of Π is $A_R^1, \ldots, A_R^n \mid\!\rightarrow A_R^{n+1}, \ldots, A_R^m$ then one can say that the conclusion of Π is the sequent $\Phi^1, \ldots, \Phi^n \mid\!\rightarrow \Phi^{n+1}, \ldots, \Phi^m$ with $\Phi_R^i = A_R^i$. One easily checks that the Φ^i's are formulas of $L_{\Omega,S,d}$.

4.7. A *structure* for the language $L_{\Omega,S,d}$ is a pair (M, R), where M is a structure for L and $R \in |\mathrm{WF}|$; if Φ is a closed formula of $L_{\Omega,S,d}$, then $(M, R) \models \Phi$ means that $M \models \Phi_R$.

4.8. Let $\mathcal{T}$ be a denumerable set of closed formulas of $L_{\Omega,S,d}$ and let (M, R) be a structure for $L_{\Omega,S,d}$; then (M, R) is a *model* of $\mathcal{T}$ iff all formulas of $\mathcal{T}$ are true in (M, R). Proofs in $\mathcal{T}+Lk_{\Omega,S,d}$ are defined exactly as in 4.4, except that proper axioms Φ_R $(\Phi \in \mathcal{T})$ are allowed in Π_R.

4.9. SOUNDNESS PROPERTY. *Assume that the closed sequent* $|\!\rightarrow \Phi$ *has a proof in* $\mathcal{T}+Lk_{\Omega,S,d}$ *and that* (M, R) *is a model of* $\mathcal{T}$; *then* $(M, R) \models \Phi$.

4.10. COMPLETENESS THEOREM. *Assume that* K *is denumerable and that* $d \in \mathrm{REP}(K, S)$; *if the closed formula* Φ *holds in all models of* $\mathcal{T}+Lk_{\Omega,S,d}$, *then it has a proof in* $\mathcal{T}+Lk_{\Omega,S,d}$; *this proof is recursive in the data.*

Proof. If $R \in |\mathrm{WF}|$, we build a preproof Π_R of $|\!\rightarrow \Phi_R$ by starting with the conclusion, and going upwards from premise to premise; it is possible to arrange a strategy in such a way that all possible rules are tried infinitely often... Assume that we have obtained the "hypothesis" $\Gamma |\!\rightarrow \Delta$, n steps above the conclusion $|\!\rightarrow \Phi_R$, and that this sequent belongs to $Lk_{(R\oplus S)+1,d'+1}$ with $d' \in \mathrm{WF}(K', R\oplus S)$, $|K'|-|K|$ finite. Among the possibilities to go upwards we have:

(i) The case of a rule $(r\mathbb{\Lambda})$, typically when $\Delta = \mathbb{\Lambda}_{a\in|R\oplus S|} A_a, \Delta'$. We then put just above $\Gamma |\!\rightarrow \Delta$ all sequents $\Gamma |\!\rightarrow A_a, \Delta$ with $a \in |R\oplus S|$, and these new "hypotheses" belong to $Lk_{(R\oplus S)+1,d'*\bar{K}a+1}$.

(ii) The case of a rule $(l\mathbb{\Lambda})$, typically when $\Gamma = \Gamma', \mathbb{\Lambda}_{a\in|R\oplus S|} A_a$. Assume that $|K| = \bigcup_p |K_p|$, with $\mathrm{card}(K_p) = p$ and assume that $(|K'|-|K|)\cup|K_n|$ is equal to $\{i_1, \ldots, i_k\}$; then we can form a finite sequence of rules $(l\mathbb{\Lambda})$ and (lC) leading from $\Gamma, A_{d'(i_1)}, \ldots, A_{d'(i_k)} |\!\rightarrow \Delta$ to $\Gamma |\!\rightarrow \Delta$.

(iii) The case of a cut: the cut formula must be taken from among subformulas of $T_R \cup \{A_R\}$ in the language $L_{(R\oplus S)+1,d'+1}$; these subformulas can be enumerated into a sequence $(B_R^0, \ldots, B_R^p, \ldots)$, and the enumeration is "functorial", i.e., $(\varrho\oplus \mathrm{id}_S)+1^{-1}(B_R^p) = B_{R_1}^p$ when

$$\varrho\otimes \mathrm{id}_S \in \mathrm{WF}^{K'}(R_1\oplus S, d_1'; R\oplus S, d');$$

also $\lambda_R^{-1}(B_R^p) = B_{R\oplus S}^p$. Then, by a big family of cuts on $B_R^0, \ldots, B_R^{n-}$ we can infer $\Gamma |\!\rightarrow \Delta$ from 2^n sequents

$$\Gamma, B_R^{i_1}, \ldots, B_R^{k^i} |\!\rightarrow B_R^{j_1}, \ldots, B_R^{j_k}, \Delta.$$

(iv) If Γ and Δ have some atomic formula B in common, or if Δ explicitly contains an axiom of $\mathcal{T}_R$, i.e., a formula B_R^p (see (iii)) which is, by definition, the qth axiom of $\mathcal{T}_R$, with $q > n$, then the process stops.

The preproofs $(\Pi_R)_{R\in|\mathrm{WF}|}$ satisfy properties 4.4(i)–(iii), except perhaps well-foundedness. Now, if for some $R \in |\mathrm{WF}|$, Π_R has an infinite branch $(\Gamma_n \mid\!\rightarrow \Delta_n)_{n\in N}$, one easily shows that the set of terms occurring in these sequent, equipped with the truth definition: $M \models A$ when $A \in \bigcup_n \Gamma_n$, $M \models \sim A$ when $A \in \bigcup_n \Delta_n$ defines a model M of $T_{R'} + \sim \Phi_{R'}$, where R' is the restriction of R to the set of all $a \in |R|$ which occur in some of these sequents. So $(M, R') \models \sim\Phi$ and is a model of $\mathscr{T}$, contradiction. ∎

5. The Ω-rule

5.1. Our task is now to reformulate our results in a more natural language, where formulas are built from the atomic ones by means of connectives and quantifiers (and similarly proofs are obtained from the axioms by means of rules).

5.2. Let Φ be a formula of $L_{\Omega,S,d}$ with $d \in \mathrm{WF}(K, S)$;

(i) If $\Phi_\varnothing$ is an atomic formula A, then $\Phi_R = A$ for all R, and we can therefore identify Φ with the atomic formula A.

(ii) If $\Phi_\varnothing$ begins with Qx, then $\Phi_R = Qx\Psi_R\theta_R$ for some $(\Psi_R)_{R\in|\mathrm{WF}|}$ and $(\theta_R)_{R\in|\mathrm{WF}|}$; it is immediate that (Ψ_R) and (θ_R) define formulas Ψ and θ of $L_{\Omega,S,d}$: so we shall represent Φ by the notation: $\Phi = Qx\Psi\theta$.

(iii) If $\Phi_\varnothing$ begins with $\bigwedge\!\!\!\bigwedge_{a(\varnothing\oplus S)+1d(i)}$ with $i \neq \bar{K}$, then $d(i) \in |S|$, so we can write without excessive abuse of notations: $\Phi_R = \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a_R$. For all $a \in |S|$ such that $aSd(i)$, the family $(\Psi^a_R)_{R\in|\mathrm{WF}|}$ defines the formula Ψ^a of $L_{\Omega,S,d*ia}$: so we shall represent Φ by the notation: $\Phi = \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a$.

(iv) If $\Phi_\varnothing$ begins with $\bigwedge\!\!\!\bigwedge_{a(\varnothing\oplus S)+1d+1(\bar{K})}$, then one can write $\Phi_R = \bigwedge\!\!\!\bigwedge_{a\in|R\oplus S|} \Psi^a_R$ (with little abuse of notation); observe that $\Psi^a_R \in L_{(R\oplus S)+1,d+1*\bar{K}a}$ and that when $\varrho \in \mathrm{WF}(R, R')$ we have $(\varrho \oplus \mathrm{id}_S) + 1^{-1}(\Psi^{\varrho(a)}_{R'}) = \Psi^a_R$. Now, take any $R \in |\mathrm{WF}|$ and define $\theta^R \in L_{\Omega,R+1\oplus S,d'}$ by $\theta^R_T = \Psi^{\bar{R}}_{T\oplus R+1}$ (d' is the extension of d to $K \oplus \{j\}$ defined by $d'(j) = \bar{R}$). Here we represent Φ by:

$$\Phi = VR \in \mathrm{WF}\,\theta^R.$$

5.3. We have so far succeeded in associating with any formula Φ of $L_{\Omega,S,d}$ an expression of that formula from other formulas (which will be styled the *subformulas* of Φ) and (generalized) connectives and quantifiers. (Observe that in the case 5.2(iv) the subformulas involved form a proper class: this means that the expressions found in 5.2 are by no means small; however, for obvious direct limit reasons, the family $(\Phi^R)_{R\in|\mathrm{WF}|}$ is completely and effectively determined by its restriction to $|\mathrm{FIN}|$.)

Two questions remain:

(Q1) Given the expressions found for Φ in 5.2, can we recover Φ (*unicity*)?

(Q2) Do all similar expressions correspond to a formula (*existence*)?

(i) In the case 5.2(i): (Q1) obviously the atomic formula A determines Φ. (Q2) any atomic formula of L leads to some Φ.

(ii) In the case 5.2(ii): (Q1) from $\Phi_R = Qx\Psi_R\theta_R$ it is clear that we can recover Φ from Ψ and θ. (Q2) given any Ψ and θ we can build $\Phi = Qx\Psi\theta$ using the equation $\Phi_R = Qx\Psi_R\theta_R$.

(iii) In the case 5.2(iii): (Q1) using $\Phi_R = \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a_R$, one can recover Φ from the family (Ψ^a). (Q2) Given such a family (Ψ^a), it is clear that we can define Φ by $\Phi_R = \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a_R$.

(iv) In the case 5.2(iv): (Q1) we use the following property of replete morphisms: if ϱ is replete, then $\varrho^{-1}(A)$ determines A uniquely. Now take $a \in |R \oplus S|$; if $a \in |R|$, define R_a by restricting R to $\{b; bRa\}$; then there is a replete morphism $\sigma \in \mathrm{REP}(R \oplus R_a + 1, R)$ such that $\sigma(\bar{R}_a) = a$, and so we get $\theta^{R_a}_R = (\sigma \oplus \mathrm{id}_S) + 1^{-1}(\Psi^a_R)$; if $a \in |S|$, define λ_R as in 4.2(ii), and let b be the element of $\lambda^{-1}_R(\{a\})$ in the first copy of S (in $R \oplus S \oplus S$); then $\Psi^b_{R\oplus S} = \lambda^{-1}_R(\Psi^a_R)$. (Q2) Given any family $(\theta^R)_{R|\mathrm{WF}|}$ such that $\theta^R \in L_{\Omega, R+1\oplus S, d'}$ there is no way of constructing a functor like Φ unless some inner solidarity is requested between the formulas θ^R. In general, when $A \in L_{\Omega,S_2,d_2}$ and $\varrho \in \mathrm{WF}^K(S_1, d_1; S_2, d_2)$ we can define $\varrho^{-1}(A)$ by

$$\varrho^{-1}(A)_R = (\mathrm{id}_R \oplus \varrho) + 1^{-1}(A_R).$$

The family (θ^R) found in 5.2(iv) enjoys the *solidarity* condition:

(SOL) *if* $\varrho \in \mathrm{WF}(R, R')$, *then* $\big((\varrho + 1) \oplus \mathrm{id}_S\big)^{-1}(\theta^{R'}) = \theta^R$.

It is not hard to see that given any family (θ^R) enjoing this property, the answer just given to (Q1) leads to a Φ such that $\Phi = \bigvee R \in \mathrm{WF}\, \theta^R$.

5.4. We have therefore obtained an alternative description[5] of the language $L_{\Omega,S,d}$ in terms of connectives and quantifiers. In fact, this description would be free from any category-theoretic considerations if we were able to formulate the condition (SOL) not for formulas viewed as functors, but for the analytic version of formulas, by means of connectives and quantifiers: this amounts to defining directly ϱ^{-1} on *analytic* formulas, and this offers neither surprise nor difficulties. Observe that the sub-

[5] "analytic" version, by opposition to the concepts of 4., which are "synthetic".

formula relation is well-founded, but that the subformulas of a given formula form in general a proper class. We can therefore perform induction on the subformula relation, such induction being much powerful than usual transfinite induction.

5.5. Now we try to find an analytic version of proofs: assume therefore that Π is a proof in $Lk_{\Omega,S,d}$.

(i) If $\Pi_\emptyset$ is an axiom $A \mid\rightarrow A$ or $\mid\rightarrow B_\emptyset$ ($B \in T$), then Π can be represented by the axiom $A \mid\rightarrow A$ or the axiom $\mid\rightarrow B$.

If $\Pi_\emptyset$ is not an axiom, let (θ) be the name of the last rule of $\Pi_\emptyset$; then, for any $R \in |\mathrm{WF}|$, the conclusion $\Gamma_R \mid\rightarrow \Delta_R$ of Π_R follows by means of the same rule (θ) from a family of premises $\Lambda_R^i \mid\rightarrow \Xi_R^i (i \in I_R)$, each of these premises being itself proved by a subproof Σ_R^i.

(ii) $(\theta) = (l1Q), (l2Q), (rC), (lC), (rW), (lW), (rE), (lE)$: for instance $(\theta) = (l2Q)$. Then $I_R = \{o\}$, $\Gamma_R = \Gamma_R'$, $Qx\Psi_R\theta_R$ and $\Lambda_R^o \mid\rightarrow \Xi_R^o$ is

$$\Gamma_R' \mid\rightarrow \theta_R[t/x], \Delta_R;$$

obviously we can say that Π has been obtained by means of the rule

$$\frac{\Gamma' \mid\rightarrow \theta[t/x], \Delta}{\Gamma', Qx\Psi\theta \mid\rightarrow \Delta}\, l1Q$$

applied to Σ^o, Σ^o being the proof in $Lk_{\Omega,S,d}$ corresponding to the family $(\Sigma_R^o)_{R\in|\mathrm{WF}|}$.

(iii) $(\theta) = (rQ), (CUT)$; then $I_R = \{o, 1\}$ and, if one defines proofs Σ^o, Σ^1 in $Lk_{\Omega,S,d}$ by the families (Σ_R^o), (Σ_R^1), one can say that Π is obtained by means of

$$\frac{\Lambda_o', \Psi \mid\rightarrow \Xi_o \quad \Lambda_1', \theta \mid\rightarrow \Xi_1}{\Lambda_o', \Lambda_1' \mid\rightarrow Qx\Psi\theta, \Xi_o, \Xi_1}\, rQ \qquad \text{or} \qquad \frac{\Lambda_o \mid\rightarrow \Phi, \Xi_o' \quad \Lambda_1', \Phi \mid\rightarrow \Xi_1}{\Lambda_o, \Lambda_1' \mid\rightarrow \Xi_o', \Xi_1}\, CUT$$

applied to Σ^o and Σ^1.

(iv) $(\theta) = (r\mathbb{M})$ and $\Delta = \underset{aSd(i)}{\mathbb{M}} \Psi^a, \Delta'$. Then $I_R = \{b; bSd(i)\}$. If $b \in I_R$ then $\Lambda_R^b = \Gamma_R$, $\Xi_R^b = \Psi_R^b, \Delta_R'$; if we define Σ^b by the family (Σ_R^b), then Σ^b is a proof of $\Gamma \mid\rightarrow \Psi^b, \Delta'$, and we can apply the rule

$$\frac{\ldots \Gamma \mid\rightarrow \Psi^b, \Delta' \ldots \text{all } bSd(i)}{\Gamma \mid\rightarrow \underset{bSd(i)}{\mathbb{M}} \Psi^b, \Delta'}\, r\mathbb{M}$$

to the family of proofs $(\Sigma^b)_{bSd(i)}$ with $\Sigma^b \in Lk_{\Omega,S,d*ib}$, to obtain Π.

(v) $(\theta) = (l\bigwedge\!\!\!\bigwedge)$ and $\Gamma = \Gamma', \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a$; then $I_R = \{o\}$, and

$$\Delta_R^o = \Gamma_R', \Psi_R^{d(j)}$$

$\Xi_R^o = \Delta_R$. If we define a proof Σ^o in $Lk_{\Omega,S,d}$ of $\Gamma', \Psi^{d(j)} \mid\!\!\rightarrow \Delta$ by the family (Σ_R^o), then it is clear that the rule

$$\frac{\Gamma', \Psi^{d(j)} \mid\!\!\rightarrow \Delta}{\Gamma', \bigwedge\!\!\!\bigwedge_{aSd(i)} \Psi^a \mid\!\!\rightarrow \Delta}\, l\bigwedge\!\!\!\bigwedge$$

enables us to pass from Σ^o to Π.

(vi) $(\theta) = (r\bigwedge\!\!\!\bigwedge)$ and $\Delta = \dot{\forall} R \in \mathrm{WF}\, \theta^R, \Delta'$; then the indexing set I_R equals $|R \oplus S|$ and, for $a \in |R|$, $\Delta_R = \Gamma_R$ and $\Xi_R^a = \Psi_R^a, \Delta_R'$. Now define Z^R by $Z_T^R = \Sigma_{T \oplus R + 1}^{\bar{R}}$; it is immediate that Z^R is a proof of $\Gamma \mid\!\!\rightarrow \theta^R, \Delta'$ in $Lk_{\Omega, R+1\oplus S, d'}$, and so we can say that Π is obtained from the family $(Z^R)_{R \in |\mathrm{WF}|}$ by applying the rule

$$\frac{\ldots\ \Gamma \mid\!\!\rightarrow \theta^R, \Delta'\ \ldots\ \text{all } R \in |\mathrm{WF}|}{\Gamma \mid\!\!\rightarrow \forall R \in \mathrm{WF}\, \theta^R, \Delta'}\, r\Omega$$

This rule is also called the *Ω-rule.*

(vii) $(\theta) = (l\bigwedge\!\!\!\bigwedge)$ and $\Gamma = \Gamma', \forall R \in \mathrm{WF}\theta^R$; then $I_R = \{o\}$ and

$$\Delta_R^o = \Gamma_R', \Psi_R^{d(i)};$$

$\Xi_R^o = \Delta_R$. We would like to make $\Psi_R^{d(i)}$ appear as something in terms of the family (functor!) $(\theta^R)_{R \in |\mathrm{WF}|}$. First observe that $d(i) \in |S|$. Given any $a \in |S|$, we can form θ^a, which is θ^{S_a}, with $S_a = S\{b \mid bSa\}$. θ^a is a formula in $L_{\Omega, S_a+1\oplus S, d}$ and if ϱ_a is the canonical morphism from $S_a + 1 \oplus S$ to S (ϱ_a is replete), then the unique solution Φ^a of $\varrho_a^{-1}(\Phi^a) = \theta^a$ is denoted by $(\theta^R)_{R \in |\mathrm{WF}|}[a]$. It is clear that

$$\Phi_R^{d(i)} = \Psi_R^{d(i)},$$

as a consequence of 4.2(iii). Hence we can obtain Π from the proof Σ^o defined by the family (Σ_R^o) by means of the rule

$$\frac{\Gamma', (\theta^R)_{R \in |\mathrm{WF}|}[d(i)] \mid\!\!\rightarrow \Delta}{\Gamma', \forall R \in \mathrm{WF}\, \theta^R \mid\!\!\rightarrow \Delta}\, l\Omega.$$

5.6. If we now ask the questions of unicity and existence for the analytic proofs in a way similar to 5.3, then we obtain an answer similar to the one found for analytic formulas. In particular, a solidarity condition,

$$\text{(SOL) } \textit{if } \varrho \in \mathrm{WF}(R, R'), \textit{ then } \big((\varrho+1)\oplus \mathrm{id}_S\big)^{-1}(Z^{R'}) = Z^R,$$

is requested in case (vi).

5.7. Finally, the Ω-rule is very close to the quantifier rules of usual logic: $(r\Omega)$ behaves like a rule $(r\forall)$, the functorial dependence being the generalization of the dependence upon a variable. More precisely, we see that *eigenvariables* are replaced in the context of the Ω-rule by the arguments R in |WF| upon which our proof functorially depends.

$(l\Omega)$ behaves like a rule $(l\forall)$: we have defined the notion of substitution of a "constant" (i.e., substitute S_a for the eigenvariable R), namely when we form $(\theta^R)_{R\in|\mathrm{WF}|}[a]$. The change of S in passing from a proof to a subproof simply reflects the fact that some eigenvariables of Π become constants in the subproof, similarly to the fact that a variable not used as an eigenvariable (i.e., not "closed") is a constant. It would be notationally convenient to represent families $(\theta^R)_{R\in|\mathrm{WF}|}$ enjoying solidarity by the notation $\theta[\xi]$, ξ being a symbol for a "variable of type WF"...

5.8. If we look at the underlying structure behind formulas and proofs, we see that these structures are trees depending functorially on R, ϱ in WF. Such trees, equipped with the familiar Kleene–Brouwer ordering, turn out to be *gerbes*, a concept which is central to the geometry of ordinals. The most typical property of gerbes is the preservation of repleteness, which has deep consequences. A typical use of gerbes in connection with Ω-logic would be to add function symbols, i.e., instead of having variables and constants (see 5.7) to have terms which are gerbes of their "variables".

5.9. With the usual β-rule, I obtained in 1979 the first full cut-elimination result for inductive definitions, with the applications mentioned in 1.5. The method goes through the iterated case, but stumbles on the case of the first recursively inaccessible ordinal, essentially because of the awkwardness of the analytic β-rule. It is likely that the Ω-rule will provide an adequate framework for a cut-elimination for the first recursively inaccessible ordinal. The treatment of Π^1_2 comprehension stumbles on the same question (the other principles seeming quite clear), and so the Ω-rule may have a large number of applications in the near future.

References

[1] Girard J.-Y., *Proof-Theory and Logical Complexity*, Bibliopolis, Napoli. The book contains a detailed bibliography on the β-rule and its applications.

[2] Girard J.-Y., *Π^1_2-Logic*, Part II: *Gerbes*, in preparation.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

PETER A. LOEB

Measure Spaces in Nonstandard Models Underlying Standard Stochastic Processes

1. Introduction

In working with stochastic processes, one often has in mind a formally finite process with infinitesimal increments or time changes. For example, the Poisson process can be thought of as a random distribution of an infinite number of unit masses into an infinite number of intervals of infinitesimal length. Brownian motion can be thought of as a random walk with infinitesimal steps. Such a cognitive experiment can be realized in a nonstandard model of the real numbers in the sense of Abraham Robinson [33]. We will show how that nonstandard cognitive experiment can be transformed into a standard probability space (still based on the nonstandard point set) that can be used as a probability space for the standard process.

Robinson's discovery is applicable to any infinite mathematical structure. A simple nonstandard model of the real numbers can be constructed by fixing a free ultrafilter Δ on the natural numbers N. Two sequences of real numbers $\{r_i\}$ and $\{s_i\}$ are equivalent if $r_i = s_i$ for all i in an element U of Δ. The equivalence classes form the nonstandard real numbers $^*R$. The constant sequence $r_i = c$ represents the standard real number c, while the sequence $r_i = i$ represents an infinite element of R and the sequence $r_i = 1/i$ represents a nonzero infinitesimal. In general, a property holds for $^*R$ if for the representing sequences it holds on some U in Δ. An element ϱ of $^*R$ is finite if $|\varrho|$ is smaller than some standard c in R and ϱ is infinitesimal if $|\varrho|$ is smaller than every positive c in R; the elements of $^*R$ that are not finite are called *infinite*. The *internal* subsets of $^*R$ correspond to equivalence classes of sequences of subsets of R; the element of $^*R$ represented by $\{r_i\}$ is in the set represented by $\{A_i\}$ if and only if

$r_i \in A_i$ for all i in some U in Δ. Not all subsets of $^*R$ are internal. Those that are not are called *external*. Some internal sets, called ***hyperfinite sets***, have all of the formal properties of finite sets. Such a set A is represented by a sequence of sets $A_i \subseteq R$ with A_i finite for all i in some U in Δ. The "internal" cardinality of A is represented by the sequence $\{\mathrm{Card}(A_i)\}$ with 0 replacing infinite cardinals in the sequence.

It is usually better to ignore any particular construction of nonstandard models and think only of the properties they satisfy. In general one starts with a set theoretic structure $V(S)$ where S is a set containing R and $V(S)$ consists of all the sets one can obtain from S in a finite number of steps using the usual operations of set theory. For example, the set of all Lebesque measurable sets is in $V(S)$ as is the set of all Borel measures on R. Let Λ be a formal language for $V(S)$; Λ contains a name for each object in $V(S)$, variables, connectives (i.e., $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftarrow$), quantifiers, brackets, and sentences formed with these symbols. Robinson [33] has shown that there is a (not unique) structure $V(^*S)$ built from a set of individuals $^*S$ with the following properties:

(1) Every name of an object in $V(S)$ names something of the same type (i.e., constructed with exactly the same operations) in $V(^*S)$. We write $^*A$ for the object in $V(^*S)$ with the same name as A in $V(S)$; A is called *standard* and $^*A$ the (*nonstandard*) *extension* of A.

(2) (Transfer Principle) Every sentence in Λ that is true for $V(S)$ is true when interpreted in $V(^*S)$; quantification, however, is over "internal" objects in $V(^*S)$.

(3) If $A \in V(S)$ is a set, then there is a "hyperfinite" set B which is a member of the extension $^*P_F(A)$ of the set of all finite subsets of A such that for each $a \in A$, $^*a \in B$. Thus B contains the extension of each standard element of A.

The extension of any individual s is usually denoted by s instead of $^*s$; one thinks of a set $A \subseteq S$ as being imbedded in $^*A$. *Internal* objects in $V(^*S)$ are those objects which are members of the extensions of standard objects; the noninternal objects in $V(^*S)$ are called *external*. The finite natural numbers N form, for example, an external set in the nonstandard natural numbers $^*N$. (If not, there would be a first infinite element of $^*N$, i.e., a first element of $^*N-N$, and thus a last element of N.) *Hyperfinite* sets are internal sets in internal one-to-one correspondence with an initial segment of $^*N$. Such sets have the formal combinatorial properties of finite sets. If ω is an infinite element of $^*N$, then the initial segment $\{n \in {}^*N\colon 1 \leqslant n \leqslant \omega\}$ of $^*N$ is a hyperfinite set containing every element of N.

The set of nonstandard real numbers $^*R$ contains infinite positive and infinite negative elements. Any other element ϱ of $^*R$ is infinitely close to a unique standard real number $r \in R$; that is, $\varrho - r$ is infinitesimal. We write $\varrho \simeq r$ and also $\varrho - r \simeq 0$ in this case; we also say that r is the *standard part* of ϱ. The standard part of ϱ is denoted by $\mathrm{st}(\varrho)$ or $^\circ\varrho$. The set of all points infinitely close to $r \in R$ is called the *monad* of r and is denoted here by $m(r)$. As an example we note that a function f is continuous on $[0, 1]$ iff for all ϱ in $^*[0, 1]$, $^*f(\varrho) \simeq f(\mathrm{st}(\varrho))$. Moreover, a set $A \subseteq R$ is compact iff for each $\varrho \in {}^*A$ there is a standard a in A with $a \simeq \varrho$ (see [33]).

One can construct a hyperfinite set X as the set of elementary outcomes in a cognitive experiment in the "nonstandard world". For example, for coin tossing one considers internal sequences of 0's and 1's of length ω where ω is an infinite element of $^*N$. The set $\mathscr{A}$ of all internal subsets of X forms an internal σ-algebra but also an algebra in the usual sense. One obtains a finitely additive probability measure P on $(X, \mathscr{A})$ by setting $P(A)$ equal to the standard part of the internal probability of A for each A in $\mathscr{A}$. At this point we need to assume that our superstructure is "denumerably comprehensive" or, equivalently, "$\aleph_1$-saturated", as will be the case if we construct it via an ultrafilter as indicated above. What this property means is that any ordinary sequence $\{A_i\colon i \in N\}$ from an internal set B is the initial segment of an internal sequence $\{A_i\colon i \in {}^*N\}$ from B. If, therefore, an ordinary sequence $\{A_i\colon i \in N\}$ from $\mathscr{A}$ is pairwise disjoint and $\bigcup_{i \in N} A_i$ equals some internal A in $\mathscr{A}$, then all but a finite number of the A_i's are empty. This follows from the fact that after extending $\{A_i\colon i \in N\}$ to $\{A_i\colon i \in {}^*N\}$, the internal set $\{a \in {}^*N\colon A \subset \bigcup_{1 \leqslant i \leqslant a} A_i\}$ contains every infinite element of $^*N$ and therefore must contain some finite element. The condition one checks to apply the Carathéodory extension theorem to $(X, \mathscr{A}, P)$ is thus vacuously satisfied, and so P has a σ-additive extension to the smallest σ-algebra Σ generated by $\mathscr{A}$. We now have a standard probability space (X, Σ, P) on the hyperfinite set X. One can use a measure space approach to develop the properties of this probability space (see [18]) or, as in [25], one can start from scratch with a functional approach. We will take the latter course.

2. A functional approach to nonstandard probability theory

Throughout the rest of this note, X will be an internal set in an $\aleph_1$-saturated enlargement of a structure containing the real numbers R. We let L denote an internal vector lattice of $^*R$-valued functions on X with $1 \in L$, and we let I denote an internal (not necessarily order continuous) positive linear functional on L with $I(1)$ finite. Here we are using the pointwise ordering for the lattice operations. In the previous section, for example, L might consist of all $\mathscr{A}$-simple functions on X and I might be the integral with respect to P. Another example is obtained by starting with a compact Hausdorff space Y and the set $C(Y)$ of all continuous real-valued functions on Y. For this example $X = {}^*Y$, $L = {}^*C(Y)$, and I is an internal positive linear functional. The Riesz representation theorem for $C(Y)$ will be a simple corollary of what follows if I is the extension of a standard or functional.

Definition 2.1. The *class of null functions* L_0 is the set of all internal and external ${}^*R$-valued functions h on X such that for any $\varepsilon > 0$ in R there is a $\varphi \in L$ with $|h| \leqslant \varphi$ and $I(\varphi) < \varepsilon$. The class L_1 consists of all real-valued functions f on X such that for some φ in L and some h in L_0, $f = \varphi + h$.

Proposition 2.2. *The sets L_0 and L_1 are vector lattices over R. If f is in L_1 with $f = \varphi + h$ for $\varphi \in L$ and $h \in L_0$, then $I(|\varphi|)$ is finite. If $\tilde{f}$ is also in L_1 with $\tilde{f} = \tilde{\varphi} + \tilde{h}$ for $\tilde{\varphi} \in L$ and $\tilde{h} \in L_0$, then $(f \vee \tilde{f}) - (\varphi \vee \tilde{\varphi}) \in L_0$ and $(f \wedge \tilde{f}) - (\varphi \wedge \tilde{\varphi}) \in L_0$. Moreover, if $f = \tilde{f}$, then $\varphi - \tilde{\varphi} \in L_0$ whence ${}^\circ I(\varphi) = {}^\circ I(\tilde{\varphi})$.*

Proof. We will show that $I(|\varphi|)$ is finite and leave the rest to the reader. Fix ψ in L with $|h| \leqslant \psi$ and $I(\psi) \leqslant 1$. Then $\varphi - \psi \leqslant f \leqslant \varphi + \psi$. Since f is real-valued, the internal set $\{n \in {}^*N: \varphi - \psi \leqslant n\}$ contains every infinite element of ${}^*N$ and thus some finite element. Similarly, $\varphi + \psi \geqslant -n$ for some n in N. It follows that $I(|\varphi|)$ is finite.

We now obtain a well defined positive linear functional J mapping L_1 into R by setting $J(f) = {}^\circ I(\varphi)$ when $f = \varphi + h \in L_1$ with $\varphi \in L$ and $h \in L_0$. It follows from $\aleph_1$-saturation that J already has the monotone convergence property.

Theorem 2.3. *If $\{f_n: n \in N\}$ is an increasing sequence in L_1 with real upper envelope F and $\sup_{n \in N} J(f_n) < +\infty$, then $F \in L_1$ and $J(F) = \lim_{n \to \infty} J(f_n)$.*

Proof. By replacing f_n with $f_n - f_1$, we may assume that each $f_n \geqslant 0$. By Proposition 2.2, we may fix $\varphi_n \in L$ and $h_n \in L_0$ for each $n \in N$ so that $f_n = \varphi_n + h_n$ and $0 \leqslant \varphi_n \leqslant \varphi_{n+1}$. By the $\aleph_1$-saturation of our enlargement, there is a $\varphi_\omega \in L$ with $\varphi_\omega \geqslant \varphi_n$ for each $n \in N$ and ${}^\circ I(\varphi_\omega) = \lim_{n \in N} {}^\circ I(\varphi_n)$. We need only show that $F - \varphi_\omega \in L_0$. Fix $\varepsilon > 0$ in R. Choose for each $n \in N$ a $\psi_n \in L$ with $|h| \leqslant \psi_n$ and $I(\psi_n) < \varepsilon/2^n$. By $\aleph_1$-saturation, we may extend the sequence $\{\psi_n\colon n \in N\}$ to an internal sequence $\{\psi_n\colon n \in {}^*N\} \subset L$. We may choose $\gamma \in {}^*N - N$ so that $\psi_n \geqslant 0$ and $I(\psi_n) < \varepsilon/2^n$ when $1 \leqslant n \leqslant \gamma$ in ${}^*N$. Setting $\psi = \sum_{1\leqslant n\leqslant\gamma} \psi_n$, we have $I(\psi) < \varepsilon$. Now for each $n \in N$.

$$\varphi_n - \psi \leqslant \varphi_n - \psi_n \leqslant \varphi_n + h_n \leqslant F \leqslant (1+\varepsilon)(\varphi_\omega + \psi),$$

so

$$(\varphi_n - \varphi_\omega) - \psi \leqslant F - \varphi_\omega \leqslant \varepsilon\varphi_\omega + (1+\varepsilon)\psi.$$

The rest is clear.

We next exhibit the close relationship between the internal lattice L and the external lattice L_1. Given an ${}^*R$-valued function g on X, we will let ${}^\circ g$ be the extended real-valued function on X defined by setting ${}^\circ g(x) = {}^\circ(g(x))$ for each x in X; ${}^\circ(g(x))$ equals $+\infty$ or $-\infty$ if $g(x)$ is infinite in ${}^*R$.

THEOREM 2.4. *A real-valued function f on X is in L_1 if and only if for each $\varepsilon > 0$ in R there exist functions ψ_1 and ψ_2 in L with $\psi_1 \leqslant f \leqslant \psi_2$ and $I(\psi_2 - \psi_1) \leqslant \varepsilon$, in which case, ${}^\circ I(\psi_1) \leqslant J(f) \leqslant {}^\circ I(\psi_1) + \varepsilon$.*

Proof. Assume that f is an arbitrary real-valued function on X for which there exists an increasing sequence $\{\psi_n\colon n \in N\} \subset L$ and a decreasing sequence $\{\tilde{\varphi}_n\colon n \in N\} \subset L$ with $\psi_n \leqslant f \leqslant \tilde{\varphi}_n$ and $I(\tilde{\varphi}_n - \psi_n) < 1/n$ for each $n \in N$. By $\aleph_1$-saturation, we may extend both sequences to ${}^*N$ and choose a $\psi_\omega \in L$ with $\psi_n \leqslant \psi_\omega \leqslant \tilde{\varphi}_n$, whence $\psi_n - \tilde{\varphi}_n \leqslant f - \psi_\omega \leqslant \tilde{\varphi}_n - \psi_n$ for each $n \in N$. It follows that $f - \psi_\omega \in L_0$ and thus $f \in L_1$. The rest is left to the reader.

PROPOSITION 2.5. *If $\varphi \in L$ with $\varphi(x)$ finite in ${}^*R$ for each $x \in X$, then ${}^\circ\varphi \in L_1$, $\varphi - {}^\circ\varphi \in L_0$,* and $J({}^\circ\varphi) = {}^\circ I(\varphi)$.

Proof. The result follows from the fact that for each $\varepsilon > 0$ in R, $|\varphi - {}^\circ\varphi| \leqslant \varepsilon$, so $\varphi - {}^\circ\varphi \in L_0$.

DEFINITION 2.6. Let M^+ denote the set of nonnegative, extended real-valued functions g on X such that for each $n \in N$, $g \wedge n \in L_1$, and set $J(g) = \sup_{n\in N} J(g \wedge n)$ for each g in M^+. Let $M = \{g\colon g \vee 0 \in M^+$ and $-g \vee 0 \in M^+\}$. For each g in M, set $J(g) = J(g \vee 0) - J(-g \vee 0)$ if at least one of the right hand values is finite. Let $\Sigma = \{A \subseteq X\colon \chi_A \in M\}$, and for each A in Σ let $\mu(A) = J(\chi_A)$.

We will show that (X, Σ, μ) is a standard measure space and that $J(g) = \int_X g\,d\mu$ for each $g \in M^+$. Note that in Definition 2.6, we may replace the truncations $g \wedge n$ with truncations $g \wedge f$ for arbitrary elements f of L_1^+ since $g \wedge f = \sup_{n \in N} g \wedge f \wedge n$.

PROPOSITION 2.7. *Fix a sequence* $\{g_n: n \in N\} \subset M^+$ *and an* $a \geqslant 0$ *in* R. *Then* g_1+g_2, ag_1, $g_1 \vee g_2$ *and* $g_1 \wedge g_2$ *are in* M^+. *Moreover,* $J(g_1+g_2) = J(g_1)+J(g_2)$, $J(ag_1) = aJ(g_1)$, *and if* $g_1 \leqslant g_2$ *then* $J(g_1) \leqslant J(g_2)$. *If* $g_n \nearrow G$ *then* $G \in M^+$ *and* $J(G) = \sup_{n \in N} J(g_n)$.

Proof. Given $n \in N$, $(g_1+g_2) \wedge n = [(g_1 \wedge n)+(g_2 \wedge n)] \wedge n \in L_1$, and for $a > 0$, $(ag_1) \wedge n = a\big(g_1 \wedge (n/a)\big) \in L_1$. The additivity of J follows from the inequality

$$J\big((g_1+g_2)\wedge n\big) = J(g_1 \wedge n)+J\big((g_1+g_2)\wedge n-(g_1\wedge n)\big) \leqslant J(g_1)+J(g_2) \leqslant J(g_1+g_2).$$

The rest is clear.

THEOREM 2.8. *The collection* Σ *is a* σ*-algebra in* X, *and* μ *is a complete, countably additive, finite measure on* (X, Σ).

THEOREM 2.9. *A nonnegative extended real-valued function* g *on* X *is* Σ*-measurable if and only if* $g \in M^+$ *in which case* $J(g) = \int_X g\,d\mu$.

Proof. Fix g in M^+ and $a > 0$ in R; let $A = \{g > a\}$. To show that the characteristic function χ_A is in L_1, we may assume that $a = 1$ since $A = \{a^{-1}g > 1\}$. Let $f = (g\wedge 2)-(g\wedge 1)$. Then $f \in L_1$ and $\chi_A = \lim_{n \in N} (1 \wedge nf) \in L_1$. The case $a = 0$ follows from Theorem 2.8, so g is Σ-measurable. The converse and the equality $J(g) = \int_X g\,d\mu$ follow from the corresponding facts for finite sums $\sum_{1 \leqslant i \leqslant n} a_i \chi_{A_i}$, $A_i \in \Sigma$.

It follows from Proposition 2.5 that if $\varphi \in L$ then $^\circ\varphi \in M$. Following Anderson [1], we call φ *S-integrable* if $I(|\varphi|)$ is finite and $J(^\circ\varphi) = {}^\circ I(\varphi)$. The reader will recognize Proposition 2.10 as an application of the usual procedure for extending integrals from bounded functions to unbounded functions. Proposition 2.11 shows that an element of M can be "lifted" to an element of L.

PROPOSITION 2.10. *A function* φ *in* L *is S-integrable if and only if for each* $\omega \in {}^*N-N$, $I(|\varphi|-|\varphi| \wedge \omega) \simeq 0$.

Proof. We may assume $\varphi \geqslant 0$. By the definition of the integral and Proposition 2.5,

$$J(^\circ\varphi) = \sup_{n \in N} J(^\circ\varphi \wedge n) = \sup_{n \in N} {}^\circ I(\varphi \wedge n).$$

Since $^{\circ}I(\varphi) = {}^{\circ}I(\varphi - \varphi\wedge n) + {}^{\circ}I(\varphi\wedge n)$ for each $n \in {}^*N$, the proposition follows.

PROPOSITION 2.11. *Given $g \geqslant 0$ in M, there is a $\varphi \geqslant 0$ in L such that for each $n \in N$, $(g\wedge n) - (\varphi\wedge n) \in L_0$, whence*

$$J(g) = \sup_{n\in N} J(g\wedge n) = \sup_{h\in N} {}^{\circ}I(\varphi\wedge n) = {}^{\circ}I(\varphi\wedge \omega)$$

for some $\omega \in {}^*N - N$.

Proof. By Theorem 2.4, we may choose sequences $\{\varphi_n \colon n \in N\}$ and $\{\psi_n \colon n \in N\}$ in L so that $0 \leqslant \varphi_n \leqslant g\wedge n \leqslant \psi_n$, $\varphi_n \leqslant \varphi_{n+1}$ and $I(\psi_n - \varphi_n) < 1/n$ for each $n \in N$. Given $k \geqslant m \geqslant n$ in N, $\psi_m\wedge n \geqslant g\wedge n \geqslant \varphi_k\wedge n \geqslant \varphi_m\wedge n$. By $\aleph_1$-saturation we may choose a $\varphi \in L$ so that for every m and n with $m \geqslant n$ in N, $\psi_m\wedge n \geqslant \varphi\wedge n \geqslant \varphi_m\wedge n$. Clearly, $g\wedge n - \varphi\wedge n \in L_0$ for each $n \in N$.

Example 2.12. Let $\mathscr{A}$ be an internal algebra (or σ-algebra) in X and let ν be an internal finitely additive (or σ-additive) measure on $(X, \mathscr{A})$. Let I be the ν-integral on the class L of internal $\mathscr{A}$-*simple functions and assume that $I(1)$ is finite in $^*R$. For each A in $\mathscr{A}$, let $\nu_0(A) = {}^{\circ}(\nu(A))$. Then ν_0 is a finitely additive measure in the ordinary sense on the algebra $\mathscr{A}$ and (X, Σ, μ) is a measure space in the ordinary sense that extends $(X, \mathscr{A}, \nu_0)$. If $B \in \Sigma$ and $\varepsilon > 0$ in R, then from the existence of functions ψ_1 and ψ_2 in L with $\psi_1 \leqslant \chi_B \leqslant \psi_2$ and $I(\psi_2 - \psi_1) < \varepsilon$, we obtain sets $A_1 = \{\psi_1 > 0\}$ and $A_2 = \{\psi_2 \geqslant 1\}$ in $\mathscr{A}$. Clearly $A_1 \subseteq B \subseteq A_2$ and $\nu(A_2 - A_1) \leqslant I(\psi_2 - \psi_1) \leqslant \varepsilon$.

Example 2.13. The set of standard real numbers can be obtained by taking the finite nonstandard rational numbers modulo the infinitesimal nonstandard rational numbers. Similarly, a standard Banach space $\hat{E}$ can be obtained from a standard Banach space E by taking the elements of $^*E$ with finite norm modulo the elements of $^*E$ with infinitesimal norm. The space $\hat{E}$ is called the *nonstandard hull* of E; $\hat{E} \neq E$ if E has infinite dimension. Extensive work on this subject has been done by W. A. J. Luxemburg, C. W. Henson and L. C. Moore, Jr. (e.g., [26, 5, 10]) who have suggested the following modification of work by Lester Helms and the author on vector integrals: Let $I(\varphi)$ take its value in an internal Banach lattice $^*E$ and $J(f)$ in the nonstandard hull $\hat{E}$. The results of this section hold provided $\hat{E}$ satisfies any condition of Theorem 4.3 of [5]. A simple, stronger condition is that when $v \geqslant u \geqslant 0$ in $^*E$ and $\|v\| \simeq \|u\|$, then $\|v - u\| \simeq 0$.

Extensions of the results of this section to the case that either $1 \notin L$ or $I(1)$ is infinite can be found in [18], [21] and [25].

3. Internal functionals on continuous functions and weak convergence

In this section we consider a set Y with a compact Hausdorff topology $\mathcal{T}$ in an enlargement of a structure containing Y and R. We assume that the enlargement is $\aleph$-saturated with $\aleph \geqslant \aleph_1$ and $\aleph \geqslant \mathrm{Card}(\mathcal{T})$. The consequences of this assumption will be noted below; the reader may consult [21] or [35] for the general definition. Since ${}^*Y$ is compact and Hausdorff, each point x in ${}^*Y$ is in the monad $m(y)$ of a unique standard point y in Y, where $m(y) = \bigcap_{y \in U \in \mathcal{T}} {}^*U$; we write $y = \mathrm{st}(x)$. With each extended real-valued function g on Y we associate the function $\tilde{g}$ on ${}^*Y$ where $\tilde{g}(x) = g(\mathrm{st}(x))$. With each subset A of Y we associate the subset $\tilde{A} = \bigcup_{y \in A} m(y)$ of ${}^*Y$, so that $\chi_{\tilde{A}} = \tilde{\chi}_A$ on ${}^*Y$. Finally, we fix an internal positive linear functional I on ${}^*C(Y)$ with $I(1)$ finite, and we apply our previous results and notation to the case that $X = {}^*Y$ and $L = {}^*C(Y)$.

PROPOSITION 3.1. *For each compact $K \subseteq Y$, $\tilde{K} \in \Sigma$ and if*

$$a_K = \inf\{{}^\circ I({}^*f) \colon f \in C(Y), \chi_K \leqslant f \leqslant 1\}$$

then $\mu(\tilde{K}) = a_K$.

Proof. It follows from $\aleph$-saturation that there is a $\varphi \in L$ with $\chi_{{}^*K} \leqslant \varphi \leqslant \chi_{\tilde{K}}$ and ${}^\circ I(\varphi) = a_K$. Given any $f \in C(Y)$ with $\chi_K \leqslant f \leqslant 1$ and given $\varepsilon > 0$ in R, we have $\varphi \leqslant \chi_{\tilde{K}} \leqslant (1+\varepsilon)^*f$. It follows that $\chi_{\tilde{K}} - \varphi \in L_0$, $\chi_{\tilde{K}} \in L_1$, and $\mu(\tilde{K}) = J(\chi_{\tilde{K}}) = {}^\circ I(\varphi) = a_K$.

THEOREM 3.2. *Let $\Sigma_Y = \{B \subseteq Y \colon \tilde{B} \in \Sigma\}$, and let $\mu_Y(B) = \mu(\tilde{B})$ for each $B \in \Sigma_Y$. Then Σ_Y is a σ-algebra in Y containing the Borel σ-algebra, and μ_Y is a complete regular measure on (Y, Σ_Y). A function g on Y is Σ_Y-measurable if and only if $\tilde{g}$ is Σ-measurable on ${}^*Y$, in which case, if $g \geqslant 0$ then $\int_Y g d\mu_Y = \int_{{}^*Y} \tilde{g} d\mu$. For each $f \in C(Y)$, $\int_Y f d\mu_Y = {}^\circ I({}^*f)$.*

Proof. To show that μ_Y is regular, choose $B \in \Sigma_Y$ and $\psi \in L$ with $0 \leqslant \psi \leqslant \chi_{\tilde{B}}$. Let $K = \{\mathrm{st}(x) \colon \psi(x) > 0\}$. By Luxemburg's Theorem 3.4.2 of [26], K is compact. Since $K \subseteq B$ and $\mu_Y(K) = \mu(\tilde{K}) \geqslant {}^\circ I(\psi)$, it follows from Theorem 2.4 that μ_Y is regular. If $f \in C(Y)$, then ${}^*f \in L, \tilde{f} = {}^\circ({}^*f) \in L_1$, and $\tilde{f} - {}^*f \in L_0$, whence $\int_Y f d\mu_Y = \int_{{}^*Y} \tilde{f} d\mu = {}^\circ I({}^*f)$. The rest is clear.

Example 3.3. If $Y = [0, 1]$ and I is the extension of the standard Riemann integral on Y, then a real-valued g on Y is Lebesgue integrable if and only if $\tilde{g} = \varphi + h$ where $\varphi \in {}^*C(Y)$ and $h \in L_0$. In this case, the Lebesgue integral of g equals the standard part of the internal Riemann integral of φ. A bounded g is Riemann integrable if and only if $°(*g) \in L_1$ (result with A. Cornea).

In [19], the author gave a nonstandard construction of representing measures for harmonic functions using the standard part map. Later, a connection with weak convergence of measures was established by the works of Anderson [1], Anderson and Rashid [3], Rashid [31], and the author [19] [22]. In brief, if the functional I of this section is the internal integral with respect to an internal Baire measure ν, then μ_Y is the standard part of ν in the weak* topology. That is, μ_Y can also be obtained by applying the Riesz representation theorem to the functional $f \to °\int_{*Y} *f\,d\nu$, $f \in C(Y)$. Anderson and Rashid's paper [3] indicated the importance of the coincidence of these two methods for obtaining μ_Y and led the author to consider the following standard corollary of the construction in [19].

Example 3.4. By a theorem first established by F. Riesz [32] but usually attributed to G. Herglotz [11], there exists for each positive harmonic function h on the unit disk D in the complex plane a representing measure ν_h on the unit circle C. Note that ν_h is both a measure on C and on the functions of the Poisson kernel $P(z, \cdot)$, $z \in C$. Let μ_x^r denote harmonic measure for x on the circle C_r of radius r; e.g., μ_0^r is normalized Lebesgue measure on C_r. One obtains ν_h as the weak* limit as $r \to 1$ of the product $h\mu_0^r$. It follows that if $\{A_i^r\}$ is a finite partition of C_r into intervals and $\delta_{y_i^r}$ is unit mass at some point y_i^r in A_i^r, then ν_h is also the weak* limit of the measures $\Sigma_i h(y_i^r)\mu_0^r(A_i^r)\,\delta_{y_i^r}$ as $r \to 1$ and the intervals A_i^r get smaller. The result continues to hold, however, if we replace the unit mass $\delta_{y_i^r}$ with unit mass δ_i^r on the function which equals the harmonic function $\mu_x^r(A_i^r)/\mu_0^r(A_i^r)$ inside C_r and equals 0 on and outside C_r. All these measures including the weak* limit ν_h are regular Borel measures on the product space $[0, +\infty]^D$. With appropriate modifications, this form of the Riesz–Herglotz theorem holds for both a Brelot harmonic space [20] and a Bauer harmonic space [24].

Extensions of this section's results to locally compact spaces can be found in [25]. If $(Y, \mathcal{T})$ is a completely regular Hausdorff space then Anderson and Rashid's results [3] on tightness and weak convergence can be formulated by extending internal functionals to $^*C(Z)$ where Z is the Stone–Čech compactification of Y. Tightness corresponds to having $\chi_{\tilde{Z}-\tilde{Y}} \in L_0$ in which case, $\tilde{Y} \in \Sigma$ so $Y \in \Sigma_Z$ and

$$\mu_Z(Y) = \sup_{\substack{K \text{ compact} \\ K \subset Y}} \mu_Z(K) = \mu_Z(Z).$$

4. Applications to probability theory

Applications of the above theory in [18] include coin tossing and the Poisson process. For coin tossing, fix $\eta \in {}^*N - N$ and let $X = \{-1, 1\}^\eta$ in Example 2.12. The algebra $\mathscr{A}$ consists of all internal subsets of X, and for each $A \in \mathscr{A}$, the internal probability assigned to A is the internal cardinality of A divided by 2^η. Let (X, Σ, P) denote the standard complete probability space derived from the internal space as in Example 2.12.

Anderson [1] constructed Brownian motion on the space (X, Σ, P) as follows: For each $\omega \in X$ and each $k \leqslant \eta$, let $\omega_k = \omega(k) \in \{-1, 1\}$. Let χ denote the hyperfinite random walk defined by setting

$$\chi(t, \omega) = 1/\eta \Big[\sum_{i=1}^{[\eta t]} \omega_i + (\eta t - [\eta t])\, \omega_{[\eta t]+1}\Big]$$

for each $t \in {}^*[0, 1]$ and each $\omega \in X$. Here, $[\eta t]$ denotes the largest element of ${}^*N$ less than or equal to ηt. Let $\beta(t, \omega) = {}^\circ\chi(t, \omega)$ for each $(t, \omega) \in [0, 1] \times \times X$. Anderson showed that β is a Brownian motion on (X, Σ, P), and for almost all $\omega \in X$, $\beta(\cdot, \omega)$ is a standard continuous function on $[0, 1]$.

Let X' denote the set of ω for which $\beta(\cdot, \omega)$ is a continuous function on $[0, 1]$. Even without compactness, the standard part map sending ω or rather $\chi(\cdot, \omega)$ onto $\beta(\cdot, \omega)$ is a measurable mapping from X' onto $C[0, 1]$ (with the sup-norm topology) and thus induces, as in Section III, a standard measure on $C[0, 1]$. Anderson showed in [1] that the measure obtained in this way is Wiener measure, and his construction gave a simple proof of Donsker's theorem for Brownian motion. Anderson also obtained a construction of the Itô integral as the standard part of an internal pathwise Stieltjes integral. For a number of applications of this construction see the works of Cutland [6], Keisler [15], Lindstrøm [16], Hoover and Perkins [12], Perkins [30], and the book of Stroyan and Bayod [34]. One immediate advantage of this approach is that the internal formula $(d\chi)^2 = dt$ takes the place of the heuristic standard formula $(d\beta)^2 = dt$, giving, among other things, an easy proof of Itô's lemma in [1].

For some of the further work in this area, the reader is referred to the following: (1) The forthcoming monograph by H. J. Keisler [15] in which a new strong existence theorem is obtained for stochastic differential equations. (2) The papers of E. Perkins [27] [28] [29] on local time in which, for example, a classical theorem of Lévy is strengthened by combining exceptional sets of measure 0 which depend on a space variable x into a single exceptional null set which works uniformly in x. (3) C. W. Henson's paper characterizing analytic sets [9]. (4) T. Lindstrøm's paper on the theorems of Prohorov, Sazanov and Gross [17]. (5) F. Wattenberg's paper on avoiding pathological sets [36]. (6) R. M. Anderson's paper on representations of measure spaces [2]. (7) L. Arkeryd's paper on the Boltzmann equation [4]. (8) T. Kamae's paper on the ergodic theorem [14]. (9) Papers on infinite particle systems and thermodynamic limits by A. E. Hurd [13] and L. L. Helms and the author [7] [8]. (10) The extensive literature in mathematical economics by a number of researchers including R. M. Anderson, D. J. Brown, H. J. Keisler, and S. Rashid. (11) The recent article by N. J. Cutland in the 1983 Journal of The London Math. Soc. (pp. 130–140) on internal controls and relaxed controls. (12) The forthcoming survey paper of E. Perkins [30] and book by K. Stroyan and J. M. Bayod [34].

Notes added in proof: (1) Leif Arkeryd has recently announced an extension of his work in [4] with a solution of the Boltzmann equation corresponding to specified periodic boundary conditions and quite general L^1 initial conditions that are appropriate to the physical setting. (2) The reader may wish to peruse in addition to [34] the recently announced book by S. Albeverio, J. E. Fenstad, R. Høegh-Krohn, and T. Lindstrøm titled "Nonstandard Methods in Stochastic Analysis and Mathematical Physics" and the recently announced introductory text by A. E. Hurd and P. A. Loeb titled "An Introduction to Nonstandard Real Analysis". Both books will be in the Series on Pure and Applied Mathematics published by Academic Press. (3) This research was supported in part by a grant from the U.S. National Science Foundation (MCS-8200494).

References

[1] Anderson R. M., A Nonstandard Representation for Brownian Motion and Itô Integration, *Israel J. Math.* **25** (1976), pp. 15–46.

[2] Anderson R. M., Star-Finite Representations of Measure Spaces, *Trans. Amer. Math. Soc.* **271** (1982), pp. 667–687.

[3] Anderson R. M. and Rashid S., A Nonstandard Characterization of Weak Convergence, *Proc. Amer. Math. Soc.* **69** (1978), pp. 327–332.

[4] Arkeryd L., A Nonstandard Approach to the Boltzmann Equation, *Arch. Rational Mech. Anal.* **77** (1981), pp. 1–10.

[5] Cozart D. and Moore L. C., Jr., The Nonstandard Hull of a Normed Riesz Space, *Duke Math. J.* **41** (1974), pp. 263–275.

[6] Cutland N. J., On the Existence of Solutions to Stochastic Differential Equations on Loeb Spaces, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **60** (1982), pp. 335–357.

[7] Helms L. L. and Loeb P. A., Applications of Nonstrandard Analysis to Spin Models, *J. Math. Anal. Appl.* **69** (1979), pp. 341–352.

[8] Helms L. L. and Loeb P. A., Bounds on the Oscillation of Spin Systems, *J. Math. Anal. Appl.* **86** (1982), pp. 493–502.
[9] Henson C. W., Analytic Sets, Baire Sets and the Standard Part Map, *Canadian J. Math.* **31** (1979), pp. 663–672.
[10] Henson C. W., and Moore L. C., Jr., Nonstandard Analysis and the Theory of Banach Spaces, to appear in: A. E. Hurd (ed.), *Nonstandard Analysis — Recent Developments*, Lecture Notes in Mathematics, vol. 983, Springer-Verlag, Berlin, Heidelberg, New York, 1983.
[11] Herglotz G., Über Potenzreihen mit positivem, reellem Teil im Einheitskreis, *Ber. Verhandl. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Klasse* **63** (1911), pp. 501–511.
[12] Hoover D. H. and Perkins E. A., Nonstandard Construction of the Stochastic Integral and Applications to Stochastic Differential Equations I and II, *Trans. Amer. Math. Soc.* **275** (1983), pp. 1–36, 37–58.
[13] Hurd A. E., Nonstandard Analysis and Lattice Statistical Mechanics: a Variational Principle, *Trans. Amer. Math. Soc.* **263** (1981), pp. 89–110.
[14] Kamae T., A Simple Proof of the Ergodic Theorem Using Nonstandard Analysis, *Israel J. Math.* **42** (1982), pp. 284–290.
[15] Keisler H. J., *An Infinitesimal Approach to Stochastic Analysis*, to appear as a Memoir of the Amer. Math. Soc.
[16] Lindstrøm T. L., Hyperfinite Stochastic Integration I, II and III, *Math. Scand.* **46** (1980), pp. 265–292, 293–314 and 315–331.
[17] Lindstrøm T. L., A Loeb-Measure Approach to Theorems of Prohorov, Sazanov and Gross, *Trans. Amer. Math. Soc.* **269** (1982), pp. 521–534.
[18] Loeb P. A., Conversion From Nonstandard to Standard Measure Spaces and Applications in Probability Theory, *Trans. Amer. Math. Soc.* **211** (1975), pp. 113–122.
[19] Loeb P. A., Applications of Nonstandard Analysis to Ideal Boundaries in Potential Theory, *Israel J. Math.* **25** (1976), pp. 154–187.
[20] Loeb P. A., A Generalization of the Riesz–Herglotz Theorem on Representing Measures, *Proc. Amer. Math. Soc.* **71** (1978), pp. 65–68.
[21] Loeb P. A., An Introduction to Nonstandard Analysis and Hyperfinite Probability Theory In: A. T. Bharucha-Reid (ed.), *Probabilistic Analysis and Related Topics*, Academic Press, New York, 1979, pp. 105–142.
[22] Loeb P. A., Weak Limits of Measures and the Standard Part Map, *Proc. Amer. Math. Soc.*, **77** (1979), pp. 128–135.
[23] Loeb P. A., A Regular Metrizable Boundary for Solutions of Elliptic and Parabolic Differential Equations, *Math. Annalen* **251** (1980), pp. 43–50.
[24] Loeb P. A., A Construction of Representing Measures for Elliptic and Parabolic Differential equations, *Math. Annalen* **260** (1982), pp. 51–56.
[25] Loeb P. A., A Functional Approach to Nonstandard Measure Theory. To appear in the *Proceedings of the S. Kakutani retirement conference* (Yale, June 1982).
[26] Luxemburg W. A. J., A General Theory of Monads, In: W. A. J. Luxemburg (ed.), *Applications of Model Theory to Algebra, Analysis, and Probability*, Holt, Rinehart and Winston, New York, 1969, pp. 18–86.
[27] Perkins E. A., A Global Intrinsic Characterization of Brownian Local Time, *Ann. Probability* **9** (1981), pp. 800–817.
[28] Perkins E. A., The Exact Hausdorff Measure of the Level Sets of Brownian Motion, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **58** (1981), pp. 373–388.

[29] Perkins E. A., Weak Invariance Principles for Local Time, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **60** (1982), pp. 437–451.

[30] Perkins E. A., Stochastic Processes and Nonstandard Analysis to appear in: A. E. Hurd ed., *Nonstandard Analysis — Recent Developments*, Lecture Notes in Mathematics, vol. 983 Springer-Verlag, Berlin, Heidelberg, New York, 1983.

[31] Rashid S., *Economies with Infinitely Many Traders*, Ph. D. dissertation, Yale University, May 1976.

[32] Riesz F., Sur certaines systèmes singuliers d'équations intégrales, *Ann. Sci. Ecole Norm. Sup.* **28** (1911), pp. 33–62.

[33] Robinson A., *Non-Standard Analysis*, North-Holland, Amsterdam, 1966.

[34] Stroyan K. D., and Bayod J. M., *Foundation of Infinitesimal Stochastic Analysis*, to appear in North-Holland Studies in Logic series.

[35] Stroyan K. D. and Luxemburg W. A. J., *Introduction to the Theory of Infinitesimals*, Series on Pure and Applied Mathematics, vol. 72, Academic Press, New York, 1976.

[36] Wattenberg F., Nonstandard Measure Theory, Avoiding Pathological Sets, *Trans. Amer. Math. Soc.* **250** (1979), pp. 257–368.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801, USA

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

RICHARD A. SHORE*

The Degrees of Unsolvability: The Ordering of Functions by Relative Computability

The basic notion in recursion theory is that of effective or algorithmic computability. We say that a function f: $N \to N$ is *recursive* or *computable* if there is a program for a Turing Machine (or equivalently for any non-trivial computer) which allows the machine to calculate $f(x)$ for each input x. The machine model for computability was introduced in Turing [30]. Many seemingly different approaches to defining effective computability had been and were later proposed by other workers including Church, Kleene, Markov and Post. All the proposed definitions however were eventually proved equivalent but it is Turing's approach that, at least from our current computer oriented view of calculability, seems the most convincing.

Once one has a formal definition of computability, one can attempt to prove that some problems have no algorithmic solutions. Thus for example we have the famous results on the algorithmic unsolvability of the word problem for groups, the halting problem for computer programs, Diophantine equations and many others. These results, of course, all show that there are non-recursive or non-computable functions. Once we know that such function exist a natural question is whether one can distinguish different levels of complexity among them. There are of course various ways to define relative complexity or computability but again the most general notion is based on Turing machines and was introduced in Post [17]: *A function f is* (*Turing*) *computable from* (*or recursive in*) *g, written $f \leqslant_T g$, iff there is some program for a* (*Turing*) *machine which, when the machine is equipped with a black box subroutine* (*called an oracle*) *which*

---

* The preparation of this paper was partially supported by a grant from the National Science Foundation.

computes $g(y)$ in a single step for any input y, allows the machine to compute $f(x)$ for any input x. It is this fundamental notion of relative recursiveness and the associated ordering on functions that will be the subject of this talk. Additional information, proofs and references can be found in Shore [27] and more completely in Lerman [14].

To be precise we must actually define an *equivalence relation* on functions; $f \equiv_T g$ when $f \leqslant_T g$ and $g \leqslant_T f$. The equivalence classes under this relation are called the *degrees of unsolvability*, *Turing degrees* or simply *degrees*. They are naturally ordered by the induced relation $\leqslant_T$. Thus it is the structure $\mathscr{D}$ of these degrees under $\leqslant_T$ which we will analyze. We will denote individual degrees by boldface letters $\mathbf{f}, \mathbf{g}, \ldots$, and their elements by lightface ones $f \in \mathbf{f}$, $g \in \mathbf{g}, \ldots$ We identify a set $A \in \boldsymbol{N}$ with its characteristic function and so may write $A \in \mathbf{a}$ as well. Thus for example we have a least degree $\mathbf{0}$ containing $\emptyset$ and all the other recursive functions.

As there are only countably many possible programs for a Turing machine, there are at most countably many functions recursive in a given f. As there are $2^{\aleph_0}$ many functions we see that there is no maximum degree. Indeed as $\mathscr{D}$ is an upper semi-lattice, it has no maximal elements either. ($\mathbf{f} \vee \mathbf{g}$ = degree of $f \oplus g$ where $(f \oplus g)(2n) = f(n)$ and $(f \oplus g)(2n+1) = g(n)$). We can in fact for any degree $\mathbf{f}$ find a natural example of a degree strictly above $\mathbf{f}$ by "relativizing" the halting problem to $\mathbf{f}$. Thus the halting problem is coded by the set $\emptyset' \equiv \{\langle x, y\rangle | \varphi_x$ halts on input $y\}$, where φ_x is the x-th possible Turing machines program. More generally given $\mathbf{f}$ we can run the programs for machine equipped with an oracle for some $f \in \mathbf{f}$ and so define $f' = \{\langle x, y\rangle | \varphi_x^f$ halts on input $y\}$. This operation, called the *jump*, is well defined on degrees and so gives us a degree $\mathbf{f}$ which by a diagonal argument is always strictly greater than $\mathbf{f}$($h(x) = \varphi_x^f(x)+1$ if $\langle x, x\rangle \in f'$ and 0 otherwise is clearly computable from f' but not from f as it cannot be φ_x^f for any x.)

Our picture of $\mathscr{D}$ now contains $\mathbf{0}$ and an increasing sequence of degrees $\mathbf{0}', \mathbf{0}'', \mathbf{0}^{(3)}, \ldots$. This sequence can be continued into the transfinite by taking infinitary joins at limit levels, e.g., $\mathbf{0}^{(\omega)} = \{\langle x, y\rangle | x \in \mathbf{0}^{(y)}\}$. Even then we are far from having exhausted $\mathscr{D}$. Our goal now is to fill out our picture of $\mathscr{D}$ by describing some of its first order and local structure. We will then apply these results to analyze some of its second order and global properties.

We begin with some results from the fundamental paper of Kleene and Post [10].

THEOREM 1. *Every countable partial ordering can be embedded in $\mathscr{D}$. (In fact they can be embedded in $\mathscr{D}$ $(\leqslant \mathbf{0}')$, the degrees below $\mathbf{0}'$ and by relativizing the proof, i.e. replacing φ_x by φ_x^f, also in $\mathscr{D}$ $[\mathbf{f}, \mathbf{f}']$ for any degree $\mathbf{f}$).*

From the post-Cohen viewpoint the proof consists of building by finite extensions countably many mutually generic functions. This enables us to embed the countable atomless Boolean algebra which is universal for countable p.o.'s in $\mathscr{D}$. That the embedding can be done below $\mathbf{0}'$ follows from the facts that it suffices to insure genericity for one-quantifier arithmetic and that such sentences can be decided recursively in $\mathbf{0}'$.

This result suffices to decide the 1-quantifier theory of $\mathscr{D}$ i.e., the truth in $\mathscr{D}$ of sentences with only existential quantifiers followed by a quantifier free matrix.

COROLLARY 2. *For every* $\mathbf{f}$ *any existential sentence is true in* $\mathscr{D}(\geqslant\mathbf{f})$, *the degrees above* $\mathbf{f}$, *iff it is consistent with the axioms for a partial ordering. Thus the* 1*-quantifier theory of* $\mathscr{D}$ $(\geqslant \mathbf{f})$ *is decidable and independent of* $\mathbf{f}$.

Once we have such an embedding result the next step up in analyzing the local structure of $\mathscr{D}$ is to ask about possible extensions of embeddings. Again a result from Kleene and Post [10] (Theorem 2.1) easily supplies some information:

THEOREM 3. *If* $P_1 \subseteq P_2$ *are finite upper semi-lattices such that no element of* $P_2 - P_1$ *is below any element of* P_1 *and* $F\colon P_1 \to \mathscr{D}$ *is an order preserving embedding, then there is an order preserving* $G\colon P_2 \to \mathscr{D}$ *which extends* F.

The obvious question raised by this result is whether the restriction that no new elements be put below any old ones is necessary. In the simplest case this asks if there is a minimal degree, that is a non-computable function f such that any non-trivial reduction in its information content produces a recursive function. This question was answered affirmatively in Spector [29].

THEOREM 4. *There is a minimal degree* $\mathbf{f} > \mathbf{0}$, *i.e., one such that* $(\mathbf{0}, \mathbf{f}) = \emptyset$. *Indeed by relativizing the proof, every degree* $\mathbf{g}$ *has a minimal cover* $\mathbf{f}$, *i.e.,* $(\mathbf{g}, \mathbf{f}) = \emptyset$.

The proof, again from a later set theoertic point of view, procedes by forcing with recursive perfect trees. As it there suffices to decide 2-quantifier sentences one can always find a minimal cover $\mathbf{f}$ of $\mathbf{g}$ below $\mathbf{g}''$. Considerable elaborations of this technique produced many difficult results describing the possible initial segments of $\mathscr{D}$. We cite three examples.

THEOREM 5 (Lachlan [11]). *Every countable distributive lattice L with 0 and 1 is isomorphic to an initial segment of $\mathscr{D}$. In fact if L is presentable recursively in* $\mathbf{f}$ *then it is isomorphic to initial segments of* $\mathscr{D}[\mathbf{f}, \mathbf{f}'']$ *and* $\mathscr{D}(\leqslant \mathbf{f}'')$.

THEOREM 6 (Lerman [13]). *Every finite lattice is isomorphic to initial segments of* $\mathscr{D}$ $(\leqslant \mathbf{f}'')$ *and* $\mathscr{D}$ $[\mathbf{f}, \mathbf{f}'']$ *for every* $\mathbf{f}$.

THEOREM 7 (Lachlan and Lebeuf [12]). *Every countable upper semilattice with* 0 *is isomorphic to an initial segment of* $\mathscr{D}(\geqslant \mathbf{f})$ *for every* $\mathbf{f}$.

Now Theorem 6 shows that the hypothesis in Theorem 3 is necessary and so together they determine which extensions of embedding are always possible. As every $\forall \ldots \forall \exists \ldots \exists$ sentence about $\mathscr{D}$ can be reduced to a finite set of such problems we can decide the 2-quantifiers theory of $\mathscr{D}(\geqslant \mathbf{f})$.

THEOREM 8 (Shore [21] and Lerman [14]). *For every* $\mathbf{f}$ *the 2-quantifier theory of* $\mathscr{D}$ $(\geqslant \mathbf{f})$ *is decidable and independent of* $\mathbf{f}$.

On the other hand Theorem 5 together with the undecidability of the theory of distributive lattices shows that for each $\mathbf{f}$, $Th\big(\mathscr{D}(\geqslant \mathbf{f})\big)$ is undecidable (Lachlan [11]). In fact we have

THEOREM 9 (Schmerl, see Lerman [14]). *For each* $\mathbf{f}$ *the 3-quantifier theory of* $\mathscr{D}(\geqslant \mathbf{f})$ *is undecidable.*

We now want to characterize the complexity of $Th(\mathscr{D})$ as was first done in Simpson [28]. The key missing ingredient is another result from Spector [29] characterizing the countable ideals of $\mathscr{D}$ as the intersections of principal ones. ($I \subseteq \mathscr{D}$ is an ideal if it is closed downward and under join.)

THEOREM 10. *For every countable ideal* $I \subseteq \mathscr{D}$ *there are degrees* $\mathbf{x}$ *and* $\mathbf{y}$, *called an exact pair for* I, *such that* $\mathrm{I} = \{\mathbf{z} \mid \mathbf{z} \leqslant \mathbf{x}$ *and* $\mathbf{z} \leqslant \mathbf{y}\}$.

The proof is by forcing with infinite-coinfinite conditions. Infinite extensions are used to code representatives f_i of the degrees $\mathbf{f}_i$ of I and finite ones guarantee that if $\mathbf{g} \leqslant \mathbf{x}, \mathbf{y}$ then $g \leqslant \bigoplus_{i \leqslant m} f_i$ for some m. The calculation of the needed genericity shows that one can find such $\mathbf{x}$ and $\mathbf{y}$ recursively in $\bigoplus_{m<\omega} (\bigoplus_{i\leqslant m} f_i)'$.

An important first order consequence of the theorem is a fact first proved in Kleene and Post [10].

THEOREM 11. *$\mathscr{D}$ is not a lattice as no exact pair for the ideal generated by any strictly ascending sequence of degrees can have a least upper bound.*

This result on exact pairs also shows that some second order quantification (over countable ideals in $\mathcal{D}$) can be expressed in a first order way in $\mathcal{D}$ (by quantifying over pairs of degrees). Together with the representation of all countable distributive lattices as initial segments of $\mathcal{D}$ given by Theorem 5, this result shows that the theory of countable distributive lattices with quantification over ideals is interpretable in the first order theory of $\mathcal{D}$. Nerode and Shore [15] then shows that quantification over symmetric irreflexive binary relations (and so over all predicates) on a countable domain is interpretable in this theory of lattices. This completes a new proof that $Th(\mathcal{D})$ is as complicated as posisble.

THEOREM 12 (Simpson [28]). *$Th^2(N) \leqslant_{1-1} Th(\mathcal{D})$, that is there is an effective procedure F for transforming sentences φ of second order arithmetic into first order ones φ^F about $\mathcal{D}$ such that $N \models \varphi$ iff $\mathcal{D} \models \varphi^F$. In fact for any* $\mathbf{f}$ *and φ, $\mathcal{D} \models \varphi^F$ if $\mathcal{D}\ (\geqslant \mathbf{f}) \models \varphi^F$ and so $Th^2(N) \leqslant_{1-1} Th(\mathcal{D} \geqslant \mathbf{f})$ as well.*

As Theorems 5 and 10 are true of many other degree orderings this argument also proves that $Th^2(N)$ is $1-1$ reducible to their theories. The applications include $1-1$, $m-1$, *tt*, *wtt* and arithmetic degrees. As all of these theories, including that of $\mathcal{D}$, are clearly $1-1$ reducible to $Th^2(N)$ they are all in fact $1-1$ equivalent to $Th^2(N)$ (Nerode and Shore [15]).

Calculations based on the local version of Theorems 5 and 10, i.e., the fact that one requires at most two jumps to find the desired degrees, allow us to apply these methods to arbitrary jump ideals $\mathcal{C} \subseteq \mathcal{D}$ (i.e., ones closed under jump).

THEOREM 13 (Nerode and Shore [16]). *If $\mathcal{C} \subseteq \mathcal{D}$ is a jump ideal then $Th(\mathcal{C})$ is 1–1 equivalent to the second order theory of $\mathbf{N}$ with set quantification restricted to those sets with degrees in $\mathcal{C}$.*

As formally the interpretation is independent of $\mathcal{C}$, we can distinguish between various jump ideals by noting differences in the corresponding theories of second order arithmetic.

COROLLARY 14 (Nerode and Shore [16]).

(1) *For every $n \geqslant 0$, $Th(\mathcal{D}_n) \neq Th(\mathcal{D})$ where $\mathcal{D}_n$ is the class of degrees of Δ_n^1 sets.*

(2) $Th(\mathcal{D}_0) \neq Th(\mathcal{D}_1)$.

(3) *If $V = L$ or* PD *holds then $Th(\mathcal{D}_n) \neq Th(\mathcal{D}_m)$ for $n \neq m$.*

Elaborations on these ideas using the methods of Shore [24] distinguish between the theories of degrees corresponding to different reducibilities.

THEOREM 15.

(1) (Shore [25]) *The theories of the Turing and truth-table degrees are distinct.*

(2) (Shore [26]) *The theories of the Turing and arithmetic degrees are distinct.*

We will now consider three types of questions about the global structure of $\mathscr{D}$ like those first systematically raised in Rogers [19] and [20].

(1) *Automorphisms.* What restrictions are there on possible automorphisms of $\mathscr{D}$?

(2) *Homogeneity.* The trend of all the results mentioned so far is that every property established for $\mathscr{D}$ is also seen to be true of $\mathscr{D}$ $(\geqslant \mathbf{f})$ for every $\mathbf{f}$ simply by relativizing the original proof. This suggested the homogeneity conjecture that $\mathscr{D}$ $(\geqslant \mathbf{f}) \cong \mathscr{D}$ for every $\mathbf{f}$.

(3) *Definability.* Which degrees or relations on degrees are definable in $\mathscr{D}$ i.e., solely in terms of the ordering of Turing reducibility?

Results on such questions for the structure $\mathscr{D}'$ augmenting $\mathscr{D}$ with the jump operator appear in Feiner [2], Yates [31] Jockusch and Solovay [9], Richter [18], Epstein [1], Jockusch and Simpson [7] and Simpson [28] among others. For $\mathscr{D}$ itself the first restriction on automorphisms appears in Nerode and Shore [16]; the homogeneity conjecture is first refuted in Shore [22] and [24]; and definability results first appear in Harrington and Shore [4]. The best results to date however are based on those in Jockusch and Shore [6] some of which we will now describe.

The relevant results center on the problem of deciding which degrees are minimal covers. Jockusch and Soare [8] showed that for every $n < \omega$ $\mathbf{0}^{(n)}$ is not a minimal cover and in fact $\mathbf{0}^{(n)} \vee \mathbf{f}$ cannot be minimal over $\mathbf{f}$ for any $\mathbf{f}$. On the other hand by Jockusch [5] every sufficiently large degree is a minimal cover while by Harrington and Kechris [3] Kleene's $\mathcal{O}$ is sufficiently large. These two results both use game theoretic (determinacy) methods. A new approach relating Δ_2^0 operators and the difference hierarchy to iterated generalizations of the jump (the REA operators) together with the appropriate completeness theorem gives a proof that in fact every degree above $\mathbf{0}^{(\omega)}$ is a minimal cover (Jockusch and Shore [6]). More crucial to our purposes is the stronger result proved there that if $\mathbf{f}$ is not arithmetic, i.e. $\forall n (\mathbf{f} \nleqslant \mathbf{0}^{(n)})$, then there is a $\mathbf{g} \geqslant \mathbf{f}$ such that $\mathbf{f} \vee \mathbf{0}^{(\omega)}$ is a minimal cover of $\mathbf{g}$. Combining this with the result of Jockusch and Soare [8] we see that the class of degrees of arithmetic sets is definable in $\mathscr{D}$.

THEOREM 16. *For every* $\mathbf{f}$ *the class of degrees above* $\mathbf{f}$ *and arithmetic in* $\mathbf{f}$ *is precisely*

$$\mathscr{C}^{\mathbf{f}} = \{\mathbf{g} \geqslant \mathbf{f} \mid \mathscr{D}(\geqslant \mathbf{f}) \vDash \exists \mathbf{h} \geqslant \mathbf{g}\ \forall \mathbf{k}\ (\mathbf{k} \vee \mathbf{h} \text{ is not a minimal cover of } \mathbf{k})\}.$$

As every countable distributive lattice arithmetic in $\mathbf{f}$ is an initial segment of $\mathscr{C}^{\mathbf{f}}$ by Theorem 5 and no others are by a simple quantifier counting argument and there are such lattices of every degree we can give a strong refutation of the homogeneity conjecture.

THEOREM 17. *If* $\mathscr{D}(\geqslant \mathbf{f}) \cong \mathscr{D}(\geqslant \mathbf{g})$ *then* $\mathbf{f}$ *and* $\mathbf{g}$ *are in the same arithmetic degree, i.e.,* $\mathbf{f} \leqslant \mathbf{g}^{(n)}$ *and* $\mathbf{g} \leqslant \mathbf{f}^{(n)}$ *for some* n.

The analysis used to prove Theorem 13 gives us even more.

THEOREM 18. *If* $Th(\mathscr{D}(\geqslant \mathbf{f})) = Th(\mathscr{D})$ *then* $\mathbf{f}$ *is arithmetic.*

We can prove this elementary equivalence version of Theorem 17 in general only if $\mathbf{g}$ is definable in second order arithmetic. Otherwise it depends on set-theoretic assumptions. It is true for every $\mathbf{g}$ if $V = L$ but not if PD holds.

Our definition of the arithmetic degrees in $\mathscr{D}$ can now be combined with general results from Nerode and Shore [16] to show that almost everything possible is definable in $\mathscr{D}$.

THEOREM 19. *Any relation* R *on degrees which is invariant under joining with an arithmetic degree is definable in* $\mathscr{D}$ *iff it is definable in second order arithmetic. Thus for example the following relations are definable in* $\mathscr{D}$:

$$\mathbf{f} \equiv_T \mathbf{g}^{(\omega)}, \quad \mathbf{f} \text{ is } \Delta_n^1 \text{ in } \mathbf{g} \text{ for } n \geqslant 0, \quad \mathbf{f} \equiv_T \mathcal{O}^{\mathbf{g}}, \quad \mathbf{f} \text{ is constructible in } \mathbf{g} \text{ and } \mathbf{f} \equiv_T \mathbf{g}^{\#}.$$

Turning finally to automorphisms of $\mathscr{D}$ we immediately note that Theorem 17 shows that if φ is an automorphism of $\mathscr{D}$ then $\varphi(\mathbf{x})$ and $\mathbf{x}$ are of the same arithmetic degree for all $\mathbf{x}$. One can however get an even stronger result by applying the methods used in the original refutation of the homogeneity conjecture (Shore [22]).

THEOREM 20. *If* $\varphi\colon \mathscr{D}(\geqslant \mathbf{f}) \to \mathscr{D}(\geqslant \mathbf{g})$ *is an isomorphism then there are* $\mathbf{h}$ *and* $\mathbf{k}$ *arithmetic in* $\mathbf{f}$ *and* $\mathbf{g}$ *respectively such that* $\varphi(\mathbf{x}) = \mathbf{x}$ *for every* $\mathbf{x} \geqslant \mathbf{f} \vee \mathbf{g}$. *In particular any automorphism of* $\mathscr{D}$ *is the identity on every degree above all the arithmetic ones.*

References

[1] Epstein R., *Degrees of Unsolvability: Structure and Theory*, Lect. Notes in Math. 759, Springer-Verlag, Berlin 1979.

[2] Feiner L., The Strong Homogeneity Conjecture, *J. Symb. Logic* **35** (1970), pp. 375–377.

[3] Harrington L. and Kechris A., A Basis Result for Σ_3^0 Sets of Reals With an Application to Minimal Covers, *Proc. Am. Math. Soc.* **53** (1975), pp. 445–448.

[4] Harrington L. and Shore R. A., Definable Degrees and Automorphisms of $\mathscr{D}$, *Bull. Am. Math. Soc.* (NS) **4** (1981), pp. 97–99.

[5] Jockusch C. G. Jr., An Application of Σ_4^0 Determinacy to the Degrees of Unsolvability, *J. Symb. Logic* **38** (1973), pp. 293–294.

[6] Jockusch C. G. Jr. and Shore R. A., Pseudo-Jump Operators II: Transfinite Iterations, Hierarchies and Minimal Covers, to appear in *J. Symb. Logic.*

[7] Jockusch C. G. Jr. and Simpson S. G., A Degree Theoretic Definition of the Ramified Analytical Hierarchy, *Ann. Math. Logic* **10** (1976), pp. 1–32.

[8] Jockusch C. G. Jr and Soare R. I., Minimal Covers and Arithmetical Sets, *Proc. Am. Math. Soc.* **25** (1970), pp. 865–859.

[9] Jockusch C. G. Jr and Solovay R. M., Fixed Points of Jump Preserving Automorphisms of Degrees, *Israel J. Math.* **26** (1977), pp. 91–94.

[10] Kleene S. C. and Post E. L., The Upper Semi-Lattice of Degrees of Unsolvability, *Ann. Math.* **59** (1954), pp. 379–407.

[11] Lachlan A. H., Distributive Initial Segments of the Degrees of Unsolvability, *Z. f. Math. Logic und Grund d. Math.* **14** (1968), pp. 547–472.

[12] Lachlan A. H. and Lebeuf R., Countable Initial Segments of the Degrees of Unsolvability, *J. Symb. Logic* **41** (1976), pp. 289–300.

[13] Lerman M. Initial Segments of the Degrees of Unsolvability, *Ann. Math.* **93** (1971), pp. 365–389.

[14] Lerman M., *The Degrees of Unsolvability*, Springer-Verlag, Berlin, 1983.

[15] Nerode A. and Shore R. A., Second Order Logic and First Order Theories of Reducibility Orderings. In: J. Barwise, J. Keisler and K. Kunen (eds.), *The Kleene Symposium*, North-Holland, Amsterdam, 1979.

[16] Nerode A. and Shore R. A., Reducibility Orderings: Theories, Definability and Automorphisms, *Ann. Math. Logic*, **18** (1980), pp. 61–89.

[17] Post E. L., Recursively Enumerable Sets of Positive Integers and their Decision Problems, *Bull. Am. Math. Soc.* **50** (1944), pp. 284–316.

[18] Richter L. J. C., On Automorphisms of the Degrees that Preserve Jumps, *Israel J. Math.* **32** (1979), pp. 27–31.

[19] Rogers H. Jr., *Theory of Recursive Functions and Effective Computability*, McGraw Hill Co., New York, 1967.

[20] Rogers H. Jr., Some Problems of Definability in Recursive Function Theory. In: Crossley J. N. (ed.), *Sets, Models and Recursion Theory*, Proc. of the Summer School in Mathematical Logic and Tenth Logic Colloquium Leicester, August-September 1965, North-Holland, Amsterdam, 1967

[21] Shore R. A., On $\forall\exists$ Sentences of α-recursion Theory. In: J. Fenstad *et. al.* (eds.), *Generalized Recursion Theory*, North-Holland, Amsterdam, 1978, pp. 331–353.

[22] Shore R. A., The Homogeneity Conjecture, *Proc. Nat. Ac. Sci.* **76**, No 9 (1979), pp. 4218–4219.

[23] Shore R. A., The Degrees of Unsolvability: Global Results. In: Lerman Schmerl and Soare (eds.), *Logic Year*, 1979-80, Springer-Verlag, Lectures Notes in Math. 859, 1891.

[24] Shore R. A., On Homogeneity and Definability in the First Order Theory of the Turing Degrees, *J. Symb. Logic.* **47** (1982), pp. 8–16.

[25] Shore R. A., The Turing and Truth-Table Degrees are not Elementarily Equivalent. In: D. van Dalen (ed.), *Logic Colloqium '80*, North-Holland, Amsterdam 1982.

[26] Shore R. A., The Arithmetic Degrees are not Elementarily Equivalent to the Turing Degrees, to appear in *Archiv f. Math. Logik und Grundlagenforschung.*

[27] Shore R. A., The Structure of the Degrees of Unsolvability. In: A. Nerode and R. A. Shore (eds.) *Proceedings of the 1982 AMS Summer Institute in Recursion Theory*, American Mathematical Society, Providence, RI, 1983.

[28] Simpson S. G., First Order Theory of the Degrees of Recursive Unsolvability, *Ann. Math.* **105**, 2 (1977), pp. 121–139.

[29] Spector C., On Degrees of Recursive Unsolvability, *Ann. Math.* **64** (1956), pp. 581–592.

[30] Turing A. M., Systems of Logic Based on Ordinals, *Proc. London Math. Soc* **45**, 2(1939), pp. 161–228.

[31] Yates C. E. M., Initial Segments and Implications for the Structure of Degrees. In: Hodges (ed.), *Conference in Mathematical Logic London 1970*, Lecture Notes in Mathematics 255, Springer-Verlag, Berlin, 1972, pp. 305–335.

CORNELL UNIVERSITY, ITHACA, N.Y. 14853

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

A. O. SLISENKO

Linguistic Considerations in Devising Effective Algorithms

In commemoration of S. Yu. Maslov

The topic of this talk concerns some computer science aspects of interosculation of two basic notions originating in mathematical logic, namely calculus and algorithm.

Here I would like to pay tribute to the late S. Yu. Maslov (who tragically died in an auto smash in the summer of 1982) for his enthusiasm to use calculi as instruments of investigation of various logical, algorithmic, biological and social phenomena, and for his contribution to this field of research.

Perhaps, it will not be out of place to mention here that both the basic notions mentioned above played a key role in classical problems of mathematical logic, e.g., the Hilbert problems related to them; and as the classical problems were exhausted more than 10 years ago, one can hardly find impressive works on algorithms or calculi produced in mathematical logic in recent years (one may even get the impression that logicians have not noticed the obvious impoverishment of the research area). On the other hand, both notions have become the theoretical foundation of computer science, where they have been enriched and developed.

But some crucial problems originating in computer science seem to be more in the spirit of mathematical logic and to need for their solution a conceptual clarification of their logical essence which can provide good algorithmization. As straightforward examples of such problems one can take computer theorem-proving or computer operation with texts in natural languages. Both problems seem to be closer to each other than we thought some time ago, and may be advanced more on the ground of clever data bases than of fast algorithms. This viewpoint is conventional in computer analysis of natural language texts but is not developed in the area of more or less general computer theorem provers. The extent of my knowledge does not permit me to remain any longer on the attractive but rather boggy soil of natural language analysis; I can afford only to touch upon one point of theorem proving.

Mathematical logic provides us with tools to represent any mathematical proof by using only a small number of very elementary steps. However, such a formal proof makes it practically impossible to understand and especially to explain what mathematician calls the "idea" or "general framework" of the proof. After the experiments with general computer theorem-provers accomplished in the sixties and many other, more theoretical attempts to formulate Gestalt mathematical considerations within the traditional notions of mathematical logic, it became clear that these means do not suffice to design effective algorithms for automatic theorem proving. Speaking about computer theorem-proving I only mean proving more or less routine mathematical theorems; such a possibility is consistent with our concept of mathematics and, what is more important, with mathematical practice.

A routine mathematical problem may look as follows. At one pole we have the problem itself (e.g., to find an explicit solution for some equation) and at the other pole we have some nontrivial mathematical results (identities, transformations, theorems). And we wish to reduce our problem to these nontrivial results according to some known ideas or some known scheme. In the simplest case such a schedule of reduction may prove to be an algorithm. In more interesting cases (with respect to this context) it resembles and may prove to be a calculus with rules "if you see a subformula of this or that type try to apply this or that transformation". So, to find a reduction under consideration is to find a kind of derivation in the calculus mentioned above. And the latter problem seems to be algorithmically not hard for routine mathematics, though I shall not insist on this assertion and prefer to treat it cautiously as a conjecture to be analysed for a large experimental material. The first obstacle I met when trying to accomplish this analysis was the absence of proper material. Routine mathematics remains outside good mathematical books or articles, even when it is a starting point for them. At the moment I am not able to give quite relevant illustrations for the considerations under discussion, and will try to outline some aspects of the situations which may probably appear within this approach.

One can easily see that within the framework described above, though speaking about mathematical reasoning, we deal not with logic in the ordinary sense but with some high-level operational (or procedural) semantics for sentences with the usual logical syntax. And this semantics, at least from a technical point of view, has, at first sight, nothing in common with traditional logical semantics — cf. Tseytin [9]. But this first impression may sometimes prove to be erroneous. Rather an extreme example is given by a programming system PRIZ developed by E. Tyugy and his group (Institute of Cybernetics, Tallin) — see [4], [10]. The system is designated for solving simple mechanical, geometrical and similar problems and permits the user to formulate them in some natural way without any explicit use of logic. The designers of PRIZ did not have in mind any formal logic when working out the system. The core of PRIZ is an algorithm which transforms an input description into a schedule for compiling a program to be executed to get a solution for the input problem. As was shown by G. Mints [13], this scheduler is an algorithm for seeking a proof of a formula (which is easily constructed from the input text) in a fragment of intuitionistic propositional logic. This result surely entailed a lot of useful theoretical and practical consequences, including drastic improvement of the speed of the scheduler. Unfortunately, purely logical considerations do not show practical ways for further development of such systems, because logic does not sustain its reasoning with good algorithmics.

Coming back to the idea of using certain calculi to find efficient algorithms in difficult situations, we can formulate the question in a general setting: how to use a calculus to describe simply solvable (say, in polynomial time) subclasses of difficult problems? The problem of good theoretical treatment of efficiently solvable subclasses of difficult problems arises not only and not so often in the automation of reasoning in the spirit of computational logic, but also and even more often in such areas of computer activity as artificial intelligence, pattern recognition, discrete optimization, automation of programming, and so on. Surely, I cannot give an all-embracing answer to the question; I shall only illustrate the idea of using calculi by an example from my paper [8].

We consider the Hamiltonian circuit problem, which is known to be NP-complete even for planar graphs [3]. (By the way, one can treat it in the spirit of theorem proving.) So, for simplicity one can keep in mind only planar graphs. In any case, graphs are assumed to be connected, may have multiple edges and labels on nodes and perhaps on edges. Now we would like to find and precisely describe some set of graphs with a lucid structure that could help us to solve the Hamiltonian circuit problem. Among the simplest structures of this kind are those that permit a decomposition reducing the initial problem to smaller ones. And to such data one can try to apply procedures of dynamic programming type. Let us fix this plan. A decomposable or hierarchical structure can be presented, for example, as a parsing in a grammar of a context-free type. It is not quite clear what is a context-free graph grammar because from the string-data point of view any such nontrivial grammar will be context-sensitive. So we would like to devise some type of grammars with weak context sensitivity which have a rather effective parsing (let it be polynomial time), which could be used for checking the property of being Hamiltonian with the help of some kind of dynamic programming.

We introduce the following type of context-free graph grammars (briefly: CFGG). A CFGG is defined by two alphabets (terminals and nonterminals), an axiom and a list of productions. The two alphabets are disjoint and used for labelling nodes of graphs. The axiom is of a fixed form: a single node labelled with a fixed nonterminal. Productions are defined by pairs of the form (A, G), where A is a nonterminal and G is a graph. Such a production permits the following substitution: take any node labelled with A, replace it by G and connect all the edges incident to the node to arbitrary nodes of G (the number of these edges remains unchanged). The language $L(\Gamma)$ determined by a CFGG Γ is defined, as usual, as all those graphs that can be derived from the axiom and contain no nonterminals.

LEMMA. *For any* CFGG Γ *there is a polynomial-time algorithm for recognizing* $L(\Gamma)$ *and for parsing the graphs of* $L(\Gamma)$.

The lemma can be proved similarly to the case of usual string context-free grammars. The crucial observation is as follows. The degree of all the graphs generated by a fixed grammar Γ is bounded (from above) by a constant. So any graph contractable by a backward derivation to a single node is determined by its "boundary edges" whose number is bounded by the same constant. For details see [8].

This lemma gives grounds for proving the following

THEOREM. *For any* CFGG Γ *the Hamiltonian circuit problem for graphs in* $L(\Gamma)$ *has polynomial time complexity.*

The main object which is analysed to find a succint representation of all the Hamiltonian circuits in a given graph G arises in the following manner. Suppose that G is planar (planarity is assumed only for lucidating the geometrical images), $G \in L(\Gamma)$ for some CFGG Γ and G is drawn on the plane. We take some parse tree for G in Γ which gives us a series of contractions shrinking G to the axiom. Let us consider an arbitrary contractable graph H of the type mentioned above, and let us look at the trace of a Hamiltonian circuit in H. In this situation we get some path covering of H. Suppose that H begins to contract in some backward derivation of G. Then the paths of the covering begin to glue but only at nodes. So, we have only to analyse some type of path coverings, and in fact their number for any contractable graph can be restricted by the number of such coverings for graphs from the right-hand sides of productions of Γ. Thus the information to be transferred from the sons of a parse tree to their father is bounded by a constant and this makes dynamic programming fairly effective — see [8].

The class of CFGG described above is rather weak. By no such grammar can one generate all the graphs or all the planar graphs or any set of graphs containing an infinite number of "rectangle grids" (though for such grids the Hamiltonian circuit problem is trivial).

The whole construction of the algorithm for the Hamiltonian circuit problem gives a hierarchy $X_1 \subseteq X_2 \subseteq \ldots$ of all the graphs, where for every class X_i the Hamiltonian circuit problem has polynomial time complexity. And every X_i is generated by a CFGG. So one can easily build a "universal" algorithm for solving the problem within this framework. This algorithm tries to find a parsing of a given graph in more and more powerful grammars, and having got a parsing it will solve the Hamiltonian circuit problem with the time complexity close to the size of the parse tree.

The languages of graphs generated by the grammars under consideration are, in a sense, stable with respect to deleting edges. If $\mathscr{K}$ is such a language, and we add to $\mathscr{K}$ any set of graphs obtained from the graphs of $\mathscr{K}$ by deleting edges (and maybe take away some old graphs), then the new set of graphs will also have a polynomial-time algorithm for the Hamiltonian circuit problem.

The same idea can be applied to other difficult, say NP-complete problems. But even within the restricted dynamic programming framework, to devise an appropriate type of grammars one needs to make some non-trivial research. As an example let us take the clique problem. Here an appropriate type of grammars looks as follows. The axiom is the same as above and any production has the form (A, G), again as above, but the corresponding substitutions are different. Namely, let x be a node labelled by A. Divide the edges incident to x into two sets, say E' and E. Connect the edges from E' to arbitrary nodes of G. Let $E = \{(x, x_1), \ldots, (x, x_m)\}$. Suppose that, in the whole graph where we replace x by G, the nodes $x_1, \ldots, x_m$ are in some complete subgraph $\mathscr{K}$ (otherwise, the situation is not admissible for the production). Then we take any set of nodes of G, say S, and form a complete graph from S and $x_1, \ldots, x_m$; it will be some $\mathscr{K}_{m+l}$, where $l = \mathrm{card}(S)$ (another way is to form a complete graph from S and $\mathscr{K}$).

New types of difficulties are created by difficult problems not concerning graphs, e.g., by the propositional satisfiability. All the propositional formulae clearly constitute some context-free language. So, on the one hand, we need a grammar generating formulae and, on the other hand, this grammar must take into account special considerations concerning some way of testing satisfiability. I shall not discuss here the details.

Up to now we have considered grammars suitable for dynamic programming. This method seems to be one of the simplest methods that can be incorporated in the grammar approach if we speak about classes of grammars. Other methods are waiting to be investigated.

To conclude this discussion I would like to note that the idea of using grammars for devising effective algorithms is not new. In a well-known book by Fu [2] one can find many examples in which, starting with a description of inputs by an ordinary context-free grammar or a similar grammar, one arrives at a fast practical algorithm. Ordinary, i.e., string generating, context-free grammars are not the only type of grammars which have been used in such situations. Various graph grammars give more interesting examples.

The problem of job scheduling under constraints is known to be NP-complete. But when the constraints form a general series parallel digraph the problem can be solved efficiently (one of the fastest known algorithms is due to Valdes, Tarjan, Lawler [12]). A general series parallel digraph is a directed acyclic graph whose transitive reduction is a minimal series parallel graph. The latter graphs are generated by the following graph grammar. The axiom of the grammar is a trivial single node graph whose node is both an input and an output node. The grammar has two dual productions: a parallel composition and a series composition. Having as premises two acyclic digraphs, one can simply put them together without any connections (parallel composition) or connect all the outputs of the first to all the inputs of the second (series composition). Though the whole grammar, including partial transitive closure, can hardly be treated as context-free, there is a linear time parsing for it [12].

Graph grammars are also used to devise effective algorithms for global data flow analysis of programs, e.g., see [1]. These grammars are rather special and as a rule have the "finite Church–Rosser" property.

On the whole the problem of polynomial-time parsing even for context-free graph grammars under traditional treatment (the class of our CFGG is much smaller) seems to be hopeless. G. Turán [11] showed that a grammar belonging to a rather restricted class of monotone node label controlled grammars can generate an NP-complete language.

What other general algorithmic ideas could be exploited within the framework of the grammar approach to devise effective algorithms? Geometrical or continuity considerations are worth attention when dealing with graphs originating in a straightforward way from geometrical problems. But here I would prefer to dwell upon one more developed method which resulted in fast string-matching algorithms and looks very linguistic. It might also give some hints what other types of grammars could be used in certain situations. The method I mean could be named the *method of identifier trees and their approximations*, though I do not insist even on calling it a method.

String-matching is, so to speak, a classical problem in computational complexity. Its formulation is very simple. For any input string of the form $U*V$, where U and V are strings in a fixed alphabet without $*$, determine whether V is a substring of U. The part of the input where an occurrence is looked for, is called a *text* (in our case it is U), and the part where the occurrence is checked is called a *pattern* (in our case it is V). The order of text and pattern in the input is essential when we speak about real-time algorithms. A real-time algorithm is, as usual, an algorithm which solves the problem for any meaningful prefix of the input (without seeing the remaining suffix part), spending on the processing of every new occurrence of a character a number of steps not greater than some absolute constant. It was noticed in the sixties that the string-matching problem cannot be solved in real-time on any Turing machines, or more generally, on any machines with polynomially accessible storage. I shall not dwell upon the history of the problem (one can find it in [7]), but only say that, after several steps made by different authors beginning from the late sixties, the complete solution was obtained by me somewhere in 1976–1977, for a detailed exposition see [7], where several related problems were also solved, including a rather difficult problem of finding in real time all the periodicities in a string (of course, in some succinct form). As for string-matching, it was shown in [7] that there is a random access machine whose registers have the length not greater than $\log n + \text{const}$, where n is the length of input, which in real time gives correct answers about the occurrences of a current pattern in the text. Here we have a rather rare case (probably the most nontrivial) when the question about the time complexity of a concrete problem has a complete theoretical solution.

The starting point of this solution is a comparatively simple concept of identifier tree which has a considerable linguistic spirit in the sense of the present context. (For some problems of string-matching type the first published explicit description of the concept seems to be due to P. Weiner [14].) For our situation we shall take the following definitions.

By $W(i)$ we denote the character at the ith place in the string W, $1 \leqslant i \leqslant |W|$, where $|W|$ is the length of W. Let $W(i) = \lambda$ for $i \leqslant 0$ and $|W| < i$, where λ is not in the alphabet of W. We denote the empty string by Λ. $W[i, j]$ is $W(i)W(i+1) \ldots W(j)$ if $i \leqslant j$ and is Λ otherwise. We assume that Λ has $|W|+1$ distinct occurrences in W; these occurrences are defined by the segmets $[1, 0]$, $[2, 1], \ldots, [|W|+1, |W|]$.

A segment $[i, j]$ is a repetition in W if $W[1, \max\{i, j\}]$ has two different occurrences in $W[i, j]$. (This property of a segment, as well as the properties defined below, does not depend on the part of W which lies to the right of the segment.) A segment $[i, j]$ is an *unextendable* (to the left) *repetition* in W if it is a repetition in $W[0, j]$, and $[i-1, j]$ is not. For example, $[1, 0]$ is a repetition in $W[0, 0]$. A segment $[i, j]$ is an *identifier* in W if it is not a repetition in W and $[i, j-1]$ is an unextendable repetition. For example, $[1, 1]$ is an identifier in any W.

We transfer the notions just introduced also onto strings in the following way: a string is an identifier in W if it has an occurrence $[i, j]$ in W which is an identifier, and so on.

Let W be a string. We define for it a rooted tree of strings, i.e., a tree whose nodes are labelled with characters of the input alphabet. Thus a certain string is ascribed to any branch starting at the root and ending at a node, say x; i.e., this string can be ascribed to x. The degree of the tree will not exceed the number of characters in the input alphabet.

When describing a tree $\mathscr{D}$ we use the following notations. Δ is the set of nodes of $\mathscr{D}$, r is its root, pd is the father link (i.e. $\mathrm{pd}(x)$ is the father of the node x), suc is the son link (it has two arguments), ch ascribes a character to any node of $\Delta \setminus \{r\}$. Certainly, if $\mathrm{pd}(x) = \mathrm{pd}(y)$ and $x \neq y$ then $\mathrm{ch}(x) \neq \mathrm{ch}(y)$. The string ascribed to a node x is denoted by $\omega(x)$, and ϖ is the inverse of ω.

Firstly we define a tree which represents exactly identifiers and unextendable repetitions in W. Then we append to it some new leaves: if Va, where a is a character, is a substring of W and the node $\varpi(V)$ is not a leaf, then we append $\varpi(Va)$ to D if it is not in $\mathscr{D}$.

For representing W we introduce the set $\Gamma = \{0, 1, \ldots, |W|\}$, $\Gamma \cap \Delta = \emptyset$. Let $\mathrm{ch}(i) = W(i)$, $\mathrm{pd}(i) = \max\{0, i-1\}$, $\mathrm{suc}(i, a) :=$ **if** $\mathrm{ch}(i+1) = a$ **then** $i+1$ **else** λ, $\omega(i) = W[1, i]$.

The structure described above has the following properties concerning identification of substrings in W:

(i.1) if V is a substring of W and U is the longest prefix of V represented in $\mathscr{D}$ and $|U| < |V|$, then $\overline{\omega}(U)$ is a leaf of $\mathscr{D}$;

(i.2) for every leaf x in $\mathscr{D}$ there exists the only node $z \in \Gamma \cup \Delta$ with minimal $|\omega(z)|$ such that any occurrence of $\omega(x)$ in W which is not a suffix of W is a suffix of some occurrence of $\omega(z)$ in W.

For any leaf x let $\mathrm{ext}(x)$ be the node z from (i.2). Then either $|\omega(x)| < |\omega(\mathrm{ext}(x))|$ or $\mathrm{ext}(x) \in \Gamma$. Now define $\mathrm{EXT}(x) = \mathrm{ext}^{(k)}(x)$ where $\mathrm{ext}^{(i)}$ is the ith iteration of ext and k is the minimal i for which $\mathrm{ext}^{(i)}(x)$ is not a leaf.

The whole construction permits checking whether a given U is a substring of W, as follows. Embed U in $\mathscr{D}$; on reaching a leaf go along EXT and continue embedding. In the case of successful embedding U is a substring of W, otherwise not.

The size of $\mathscr{D}$ can easily be shown to be linear. The updating of $\mathscr{D}$ can also be done in linear time with moderate effort. Besides $\mathscr{D}$ and EXT we shall build a function len such that $\mathrm{len}(x) = |\omega(x)|$ and link-functions ins and loc on $\Gamma \cup \Delta$: ins is a node $z \in \Delta$ such that $z \neq x$, $\omega(z)$ is a suffix of $\omega(x)$, the longest one among the suffixes of this type; $\mathrm{loc}(x) \in \Gamma$ and $\mathrm{loc}(x)$ is the end point of the first occurrence of $\omega(x)$ in W. After the construction of a node x representing an identifier, the link ins permits us to find the node representing the unextendable repetition which is a suffix of this identifier. This is done by calculating $z_i = \mathrm{ins}^{(i)}(\mathrm{pd}(x))$ up to getting z_i such that $\mathrm{suc}(z_i, \mathrm{ch}(x)) \neq \lambda$. For further details see [7].

An important feature of the construction outlined above is an interaction of two trees, namely of $\mathscr{D}$ which can be defined by $(\Delta, \tau, \mathrm{pd})$ and of $(\Delta, \tau, \mathrm{ins})$. The first one represents substrings (including identifiers), the second — their relations. The first one corresponds to a kind of grammar, the second one sustains the exploiting and updating of the first one.

The ideas sketched above do not suffice to construct a real-time algorithm even for string-matching, not to speak about more difficult problems, e.g., the problem of finding a nontrivial periodicity. One can hardly build a full identifier tree in real time. Luckily, some approximation to an identifier tree have proved to be sufficient to get real-time algorithms for the problems mentioned above. For example, in the case of real-time string-matching one needs, in some sense, only identifiers and unextendable repetitions of the length of the pattern to be represented correctly. And the length of pattern grows on slowly enough. One can easily make our identifier tree correct (apart from rather complicated technical details) for well-separated identifiers and repetitions using a kind of recursive doubling. But in a situation with strongly overlapping strings we stumble upon some essential difficulties. In [7] they are overcome by extending identifier trees to trees containing also periodical substrings of the input string. And for these periodicities the new extended tree has some special structure of mutual embeddings of the periodicities.

A similar idea of treating overlapping substrings by using periodicities arising when two occurrences of the same string strongly overlap was used by me in [5], where I solved the problem of real-time palindrome recognition on multihead Turing machines, having disproved the widely spread conjecture about the impossibility of real-time multihead Turing machines for palindrome recognition.

Coming back to the idea of identifier tree one can see that taken formally, it is applicable in various situations, cf. [6]. But such straightforward generalization does not usually lead to efficient algorithms. And here several ways to apply linguistic considerations in order to get good algorithms remain open. Sometimes even a superficial analysis of bad behaviour of identifier trees shows for what type of inputs the trees can be made more efficient, and the type of inputs can be expressed in a grammar form. From the theoretical point of view another possibility is more attractive. In some cases it is evident that an identifier tree can be compressed by using some compact representation of more or less "regular" parts of the input structure; in the case of string-matching such "regular" parts were periodicities, though there they were used not for compression. In more complicated structures, e.g., plane pictures, some kinds of grammars seem to be needed within this framework. But at the moment we are at the very beginning of investigation along these ways, and chances for success remain unclear.

References

[1] Farrow R., Kennedy K., and Zucconi L., Graph grammars and global program data flow analysis, *17-th Annu. Symp. Found. Comput. Sci., Houston, Texas*, 1976, pp. 42–56.

[2] Fu K. S., *Syntactic pattern recognition and applications*, Prentice Hall, Englewood, Cliffs, N. Y., 1982.

[3] Garey M. R., Johnson D. S., and Tarjan R. E., The planar Hamiltonian circuit problem is NP-complete, *SIAM J. Comput.* **5**, 4 (1976), pp. 704–714.

[4] Кахро М. И., Калья А. П., Тыугу Э. Х., *Инструментальная система программирования ЕС ЭВМ* (*ПРИЗ*), Финансы и статистика, М., 1981, 158 с.

[5] Слисенко А. О., Распознавание предиката симметрии многоголовчатыми машинами Тьюринга со входом, *Тр. Мат. ин-та АН СССР* **129** (1973), с. 30–202 (English transl. in *Proc. Steklov Inst. Math.* **129** (1973), pp. 25–208).

[6] Slisenko A. O., Computational complexity of string and graph identification, *Lect. Notes Comput. Sci.* **74** (1979), pp. 182–190.
[7] Слисенко А. О., Поиск периодичностей и идентификация подслов в реальное время, *Зап. научн. семинаров ЛОМИ* **105** (1981), pp. 62–173.
[8] Slisenko A. O., Context-free grammars as a tool for describing polynomial-time subclasses of hard problems, *Inform. Process. Lett.* **14** (1982), pp. 52–56.
[9] Tseytin G. S., From logicism to proceduralism, *Lect. Notes Comput. Sci.* **122** (1981), pp. 390–396.
[10] Tyugy E. H., The structural synthesis of programs, *Lect. Notes Comput. Sci.* **122** (1981), pp. 290–303.
[11] Turán G., *On the complexity of graph grammars*, Preprint, Automata Theory Research Group, Hungarian Acad. Sci., 1982, 18 pp.
[12] Valdes J., Tarjan R. E., and Lawler E. L. The recognition of series parallel digraphs, *11-th Annu. ACM Symp. Theory Comput.*, Atlanta, Georgia, 1979, pp. 1–12.
[13] Волож Б. М., Мацкин М. Б., Минц Г. Е., Тыугу Э. Х., Система ПРИЗ и исчисление высказываний, *Кибернетика* **6** (1982), pp. 63–70.
[14] Weiner P., Linear pattern matching algorithms, *IEEE 14-th Annu. Symp. Switch. and Automata Theory*, Iowa, 1973, pp. 1–11.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

B. I. ZIL'BER

The Structure of Models of Uncountably Categorical Theories

0. Introduction

The natural notion of categoricity, as it was discovered in the 1930's, is degenerate for first order languages, since only a finite structure can be described up to isomorphism by its first order theory. This has led to a new notion of categoricity.

A theory is said to be *categorical in a power* if it has a model of this power which is unique up to isomorphism. Morley has proved, answering the well-known question of Łoś, that for the spectrum of the infinite powers in which a given first order theory is categorical, the only possibilities are:

(i) all infinite powers,

(ii) all uncountable powers,

(iii) $\aleph_0$,

(iv) empty.

Theories having (i) or (ii) as the categoricity spectrum are called *uncountably categorical. Totally categorical* theories are those corresponding to (i) only. The same terms are used for models of such theories.

The countable $\aleph_0$-categorical structures are characterized by the Ryll-Nardzewski–Svenonius–Engeler theorem as those structures M for which there are only finitely many orbits under the action of $\operatorname{Aut} M$ on M^n, for every n. This is in fact the only general theorem known in this case and there are no grounds for expecting a good general theory.

We have an entirely different situation in the uncountably categorical case. Natural examples here are an algebraically closed field, a simple algebraic group over an algebraically closed field [19], a vector space over a (countable) division ring (with a unary operation for each scalar), the natural numbers with the successor operation.

These examples are very typical as present day theory shows.

1. Structural categoricity theory

Baldwin and Lachlan [2] were the first to demonstrate the role of strongly minimal subsets of uncountably categorical structures considered together with an algebraic closure operator.

An infinite definable subset of a structure M is called *strongly minimal* if it cannot be divided into two infinite parts by a definable subset of any elementary extension of M.

The algebraic closure $\operatorname{acl} X$ of a subset X of a structure M is the union of all finite subsets of M that are definable by means of parameters from X.

The *algebraic closure* equips a strongly minimal set with a dependence relation in the sense of Van der Waerden's "Moderne Algebra", which generalizes both the notion of linear dependence in vector spaces and the algebraic dependence of elements in fields. Such dependence systems are also called pregeometries or geometries, provided that every singleton is closed and $\operatorname{acl} \emptyset = \emptyset$ [1].

Also, Baldwin noted that when the whole structure M is connected "rigidly" with a strongly minimal subset S (i.e., $M = \operatorname{acl} S$; M is then called *almost strongly minimal*), many properties of M are determined by properties of S. Thus, a way has been outlined how to deal separately with global properties of uncountably categorical structures and their local properties, namely properties of their strongly minimal subsets.

2. Global properties

The essence of what we know about global properties of uncountably categorical structures is contained in the Theorem 2.1 below, whose implicit proof can be found in [22], [18].

Throughout this paper by *X-definable sets in M* we mean the sets of the form N/E where N is a subset of M^n and E is an equivalence relation on N and both are definable in M by means of parameters from X. An *X-definable structure in M* is an X-definable set with X-definable relations, or a structure isomorphic to it. An *X-atom* is an X-definable set which is minimal (with respect to inclusion). There is a natural construction which attaches to M any set X-definable in M and which preserves categoricity in power. So we work in a large structure including M and any given X-definable set in M.

THEOREM 2.1 (The Ladder Theorem). *For every uncountably categorical structure M there is a finite sequence $M_0, \ldots, M_n$ of M-definable sets such that*

(i) *M_0 is strongly minimal, $M_n = M$;*

(ii) *for every $k < n$ and an M_k-atom $A \subseteq M_{k+1}$ the group* $\mathrm{Gal}(A/M_k)$ *of all M_k-elementary permutations of A is M_k-definable. The action of* $\mathrm{Gal}(A/M_k)$ *on A is definable by means of parameters from $A \cup M_k$.*

Obviously, if all the groups $\mathrm{Gal}(A/M_k)$ are finite, then $M_{k+1} \subseteq \mathrm{acl}\, M_k$. Hence, if M is not almost strongly minimal, then some of those groups are infinite. If $\mathrm{Gal}(A/M_k)$ is abelian, then the group action on A is definable by means of parameters from M_k only.

A positive answer to the following question would simplify a great deal in the categoricity theory.

PROBLEM A. *Is any uncountably categorical structure a projection of an almost strongly minimal one?*

3. Local properties

Associate with a strongly minimal structure S and its subset C a *geometry* whose points are $\mathrm{acl}(C, a)$ and whose n-dimensional subspaces are $\mathrm{acl}(C, a_1, \ldots, a_{n+1})$ where $a \notin \mathrm{acl}(C)$ and $a_1, \ldots, a_{n+1}$ are algebraically independent over C. We call S *projective* if the geometry associated with S over $\emptyset$ is isomorphic to the geometry of a projective space over a division ring. We call S *locally projective* if the geometry associated with S over any non-algebraic element is projective. We call S *disintegrated* if $\mathrm{acl}(X \cup Y) = \mathrm{acl}\, X \cup \mathrm{acl}\, Y$ for any $X, Y \subseteq S$.

Lachlan introduced in [7] a notion of a pseudoplane that turned out to be of great significance in the structural theory. A *pseudoplane* is an incidence system of "points" and "lines" satisfying the following: every point (respectively, line) is incident to an infinite set of lines (points); two different lines (points) are incident in common to a finite number of points (lines).

THEOREM 3.1 (Trichotomy Theorem). *For every uncountably categorical structure M one and only one of the following holds:*

(1) *in every strongly minimal structure definable in M, a pseudoplane is definable;*

(2) *every strongly minimal structure definable in M is locally projective;*

(3) *every strongly minimal structure definable in M is disintegrated.*

For the totally categorical case this theorem was proved in [24]. It can be proved that (1) is equivalent to the definability of a pseudoplane in M. In [21] it is shown that the pseudoplanes in (1) are in fact uncountably categorical.

Each of those three types of uncountably categorical structures needs a separate treatment. As all known structures satisfying (1) arise from algebraically closed fields, we shall call them *field-like* structures. Typical examples of (2) are uncountably categorical modules studied in [17]. The structures satisfying (2) will be called *module-like*. The structures satisfying (3) will be said to be of *disintegrated type*.

4. Field-like structures

The main hypothesis about this class is:

CONJECTURE B. *Every uncountably categorical pseudoplane is definable in an algebraically closed field, and the field is definable in the pseudoplane.*

A trivial illustration of the conjecture is an affine plane over an algebraically closed field. Much more interesting is the following

Example. Let P be a k-dimensional algebraic variety over an algebraically closed field F, $k \geqslant 2$. It is well known that algebraic curves on P of a given degree can be coded by points of some algebraic set. Let $n \geqslant k$ and let L be an n-dimensional algebraic set of irreducible curves on P. It can be shown that in the incidence system (P, L) an uncountably categorical pseudoplane is definable.

It would be interesting to verify Conjecture B for this pseudoplane and in particular to show that in the incidence system (P, L) the field F is definable.

The problem stated by Conjecture B seems to be very difficult. Even the following special case is still unsolved.

CONJECTURE C. *Every uncountably categorical affine plane is desarguesian and hence is an affine plane over an algebraically closed field.*

Conjecture B, if proved, will clarify a great deal.

The following is of much interest from the point of view of model-theoretic algebra, and is significant in connection with the Ladder Theorem.

CONJECTURE D. *Any uncountably categorical simple group is an algebraic group over an algebraically closed field.*

Conjecture D was stated independently by the present author [19] (in question form) and by Cherlin [4]. Cherlin proved Conjecture C in the special case of groups of small Morley rank. One can prove:

THEOREM 4.1. *Conjecture* B *implies that any uncountably categorical simple group is definable in an algebraically closed field.*

The following theorems are also known:

THEOREM 4.2 (van den Dries [6]). *A group definable in an algebraically closed field of characteristic* 0 *is algebraic over the field.*

THEOREM 4.3 (Thomas [14]). *A locally finite simple uncountably categorical group is algebraic over an algebraically closed locally finite field.*

We conclude the section with the following theorem, which will be discussed later.

THEOREM 4.4 *There exists no totally categorical pseudoplane. In other words, no totally categorical structure is field-like.*

5. Module-like structures

THEOREM 5.1. (i) *Any group G definable in a module-like structure M is abelian-by-finite.*

(ii) *If an infinite G has no proper infinite subgroups that are definable in M, then G is abelian and strongly minimal in M.*

The theorem is proved by a direct construction of a pseudoplane in G if either of the above assertions is false.

Theorem 5.1 implies the following specification of the Ladder Theorem:

THEOREM 5.2. *For every uncountably categorical module-like structure M there is a finite sequence of sets $N_0, \ldots, N_k$ definable in M and such that*

(i) N_0 *is strongly minimal,* $\operatorname{acl} N_k \supseteq M$;

(ii) $\max\limits_{a \in N_{i+1}} \operatorname{rank}(a, N_i) = 1$ *for every* $i < k$.

In other words, in the module-like case the ladder can be constructed with steps of height 1.

Another immediate consequence of Theorem 5.1 combined with Theorem 4.4 is the well-known Baur–Cherlin–Macintyre theorem [3].

THEOREM 5.3. *A totally categorical group is abelian-by-finite.*

Theorem 4.4 on the non-existence of totally categorical pseudoplanes combined with the Trichotomy Theorem and some fairly simple combinatorical arguments implies:

THEOREM 5.4 (Classification Theorem). *A geometry associated with a strongly minimal set in a totally categorical structure is either affine or projective of infinite dimension over a finite field, or else an infinite set with the trivial closure operator.*

The theorem implies the non-existence of totally categorical pseudoplanes modulo the Trichotomy Theorem. A sketch of the proof of Theorem 5.4 is to be found in [25]. The complete proof will appear in Siberian Mathematical Journal.

Cherlin [5] has obtained the Classification Theorem independently as a consequence of the classification of all finite doubly transitive groups (which is based on the classification of all finite simple groups). The result follows from the observation that the automorphism group of any finite subspace of the geometry here considered is a doubly transitive finite group. We proved Theorem 5.4 studying pseudoplanes by model-theoretic methods and using no other deep theory. We hope that this method can be fruitful in the investigations of groups with some special transitivity conditions, including the cases in which the finite groups classification is inapplicable.

The main tool in our proof of Theorem 4.4, which is the key theorem in our situation, is

THEOREM 5.5 (Polynomial Theorem). *Let F be a definable subset and S a strongly minimal subset of an $\aleph_0$-categorical structure and $F \subseteq \operatorname{acl} S$. Then there exist a polynomial $P_F(x)$ over the rationals and a natural number n_0 such that, for every finite algebraically closed subset X, if $n = |S \cap X| \geqslant n_0$ then $|F \cap X| = P_F(n)$. The degree of P_F is equal to the Morley rank of F.*

Thus, to some extent, the polynomial P_F can be regarded as the "number of elements" of F. As we can now operate with the "number of elements" of definable subsets of a totally categorical pseudoplane, we apply the combinatorial methods traditionally used for finite incidence systems. This, in the final analysis, leads to a contradiction.

On the other hand, the correspondence $F \to P_F$ is a kind of a rank function ranging in the ring of polynomials. Note also that the classical Morley rank is a derivative of this rank. A rank of such a kind can be considered in a more general case. This works in the following situation.

The *definable closure* $\operatorname{dcl} X$ of a subset X of a structure is the union of all X-definable singletons. Obviously, $\operatorname{dcl} X \subseteq \operatorname{acl} X$.

We say that a structure M is *almost rationally equivalent* to a structure N if $M - \operatorname{dcl}\emptyset = N - \operatorname{dcl}\emptyset$ and the set of all relations on $M - \operatorname{dcl}\emptyset$ definable in M without using parameters coincides with that of N.

THEOREM 5.6. *A strongly minimal structure in which algebraic and definable closures coincide is almost rationally equivalent to one of the following structures:*

(i) *a vector space V over a division ring with a subspace V_0 of constants;*

(ii) *an affine space associated with a vector space V and expanded by the translations from V_0;*

(iii) *a set with a group of unary invertible operations each having no fixed point.*

Marczewski [9] introduced algebras admitting a special dependence relation. Later, Narkiewicz [10] extended this class and introduced v^*-algebras.

A structure in a language without predicate symbols is called a v^*-*algebra* if it satisfies the following:

For any terms f, g and independent elements $a_1, \ldots, a_n$ (which means that no a_i belongs to the subalgebra generated by all the others) we have $f(a_1, \ldots, a_n) = g(a_1, \ldots, a_n)$ iff $f = g$ is an identity.

Urbanik described all such algebras of dimension greater than 2 as the algebras (i)–(iii) of Theorem 5.6 [15]. Urbanik's Theorem for infinite dimension was essentially exploited in model theory to describe equational, Horn and universal Horn theories categorical in power (see the survey [11]), since strongly minimal subsets of the models of these theories are v^*-algebras. Applying the same methods as in the proof of Theorem 5.6 we have obtained a new proof of Urbanik's Theorem for v^*-algebras of dimension greater than 2 including the case of finite algebras.

6. Structures of disintegrated type

The Ladder Theorem in this case collapses to the following.

THEOREM 6.1. *For any structure M of disintegrated type no infinite group is definable in M, hence M is almost strongly minimal.*

The structure of strongly minimal disintegrated sets is described only in the totally categorical case.

THEOREM 6.2. *A disintegrated strongly minimal totally categorical structure is definable by means of parameters in a trivial structure (a set without any structure).*

Cherlin has recently obtained a more explicit description of such structures.

7. Finite axiomatizability problem

Most of the structural theory above was developed in an attempt to solve the finite axiomatizability problem for totally categorical structures. The finite axiomatizability problem for uncountably categorical theories was posed, as far as we know, in Vaught's survey [16]. The problem divides into two parts:

(1) Does there exist an uncountably categorical finitely axiomatizable theory which is not $\aleph_0$-categorical?

(2) Does there exist a complete totally categorical finitely axiomatizable theory?

Peretjat'kin answered "yes" to (1), constructing an example [12].

The answer to (2) is "no", which results immediately from the following

THEOREM 7.1. *Any sentence φ which is true in a totally categorical structure M is true in some finite substructure of M.*

Since the proof of Theorem 7.1 had several iterations, it is worthwhile to make some historical comments.

The first step to the theorem was made by Makowsky [8], who proved Theorem 7.1 for almost strongly minimal totally categorical M. He observed that in this case any substructure of M with the underlying set of the form $\operatorname{acl} X$ is a finite model of φ for any finite sufficiently large X.

In 1975 I gave a proof that Theorem 5.1 (ii) for totally categorical M implies Theorem 7.1 (published in [18]). In 1977 it was proved that Theorem 4.4 implies a version of the Classification Theorem 5.4 and Theorem 5.4 implies Theorem 5.1 (ii). The result was announced in [20], the complete proof was published in [21]. Thus in 1977 it was discovered that

$$4.4 \Rightarrow 5.4 \Rightarrow 5.1(\text{ii}) \Rightarrow 7.1.$$

In 1979 during my six-months' stay at the Wrocław University I succeeded in proving Theorem 7.1. The proof was based on a more conceptual version of [18]. In particular, a notion of an envelope $E(X)$ of a subset X was introduced, being already implicit in that paper. The properties of $E(X)$ are similar to those of both $\operatorname{acl} X$ and the elementary substructure prime over X. In particular, for sufficiently large finite X the envelope $E(X)$ is a finite model of a given φ. The key new point of the proof was the Polynomial Theorem.

The proof was published in [22] and its Russian version in [23]. Unfortunately, the published proof contained an error caused by an inaccuracy in the formulation and applications of the Polynomial Theorem. (It was pointed out to me by Cherlin, in the autumn of 1980.) The error could easily be corrected by a direct application of Classification Theorem 5.4, which was known to Cherlin and me by the time. Cherlin, Harrington and Lachlan [5] applied the Classification Theorem and the notion of envelope to give a new proof of Theorem 7.1 in the more general situation of $\aleph_0$-categorical $\aleph_0$-stable structures. (The paper also contains other strong results on such structures based on the Classification Theorem.)

However, there was another, more direct way to correct the proof, preserving the whole scheme of [22]. This way was to specify the Polynomial Theorem for the case where F is an abelian group. This was done in [24] by using a deep number-theoretic result about exponential diophantine equations [13]. In fact, the corrected proof is the first version of my proof of 1979; it was clear to me at the time that for the case where F is an abelian group the Polyniomal Theorem could be specified, and Professor Narkiewicz kindly informed me of the results of [13]. Unfortunately, subsequent "simplifications" of the proof reduced it to an incorrect one.

References

[1] Aigner M., *Combinatorial Theory*, Berlin–Heidelberg–New York, 1979.

[2] Baldwin J. T. and Lachlan A. H., On Strongly Minimal Sets, *J. Symb. Log.* **36** (1971), pp. 79–96.

[3] Baur W., Cherlin G., and Macintyre A., Totally Categorical Groups and Rings, *J. Algebra* **57** (1979), pp. 407–440.

[4] Cherlin G., Groups of Small Morley Rank, *Ann. Math. Log.* **17** (1979), pp. 1–28.

[5] Cherlin G., Harrington L., and Lachlan A. H., *$\aleph_0$-Categorical $\aleph_0$-Stable Structures*, preprint, 1981.

[6] van den Dries L., Definable Groups in Characteristics 0 Are Algebraic Groups, *Abstr. AMS* **3** (1982), p. 412.

[7] Lachlan A. H., Two Conjectures Regarding the Stability of ω-Categorical Theories, *Fund. Math.* **81** (1973/74), pp. 133–145.

[8] Makowsky J. A., On Some Conjectures Connected with Complete Sentences, *Fund. Math.* **81** (1973/74), pp. 193–202.

[9] Marczewski E., Independence in Some Abstract Algebras, *Bull. Ac. Pol. Sci.* **7** (1959), pp. 611–616.

[10] Narkiewicz W., Independence in a Certain Class of Abstract Algebras, *Fund. Math.* **50** (1962), pp. 333–340.

[11] Палютин Е. А., Спектр и структура моделей полных теорий. В кн.: *Справочная книга по математической логике,* Часть I: *Теория моделей,* Москва, 1982.

[12] Перетятькин М. Г., Пример $\aleph_1$-категоричной полной конечно аксиоматизируемой теории, *Алг. и лог.* **19** (1980), 314–347.

[13] Schinzel A. and Tijdeman R., On the Equation $y^m = P(x)$, *Acta Arithm.* **31** (1976), pp. 197–204.

[14] Thomas S., Model Theory of Locally Finite Groups, *Abstr. AMS* **3** (1982), p. 281.

[15] Urbanik K., A Representation Theorem for v^*-Algebras, *Fund. Math.* **52** (1963), pp. 291–317.

[16] Vaught R., Models of Complete Theories, *Bull. AMS* **69** (1963), pp. 299–313.

[17] Ziegler M., *The Model Theory of Modules,* preprint, 1981.

[18] Зильбер Б. И., *Строение моделей категоричных теорий и проблема конечной аксиоматизируемости,* Деп. № 2800–77, Кемерово, 1977.

[19] Зильбер Б. И., Группы и кольца, теории которых категоричны, *Fund. Math.* **95** (1977), pp. 173-188.

[20] Зильбер Б. И., Комбинаторные проблемы в теории категоричности. В кн.: *Математика и механика, тезисы докладов шестой Казахстанской межвузовской научной конференции,* Часть I: *Математика* (ред. А. Т. Лукьянов), КазГУ, Алма-Ата, 1977.

[21] Зильбер Б. И., Сильно минимальные счетно категоричные теории, *Сиб. мат. журн.* **21** (2) (1980), 98–112.

[22] Zil'ber B. I., Totally Categorical Theories: Structural Properties and the Non-Finite Axiomatizability. In: Pacholski *et al.* (eds.), *Model Theory of Algebra and Arithmetic,* Lecture Notes Math. **834**, 1980, pp. 381–410.

[23] Зильбер Б. И., Решение проблемы конечной аксиоматизируемости для теорий, категоричных во всех бесконечных мощностях. В кн.: *Теория моделей и её применения* (ред. Б. С. Байжанов), КазГУ, Алма-Ата, 1980.

[24] Зильбер Б. И., О проблеме конечной аксиоматизируемости для теорий, категоричных во всех бесконечных мощностях. В кн.: *Исследования по теоретическому программированию* (ред. Б. С. Байжанов), Алма-Ата, 1981, 69–75.

[25] Зильбер Б. И., Тотально категоричные структуры и комбинаторные геометрии, *ДАН СССР* **259** (1981), 1039–1041.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

ROBERT L. GRIESS, JR.

The Sporadic Simple Groups and Construction of the Monster

I. The finite simple groups

1. Introduction. At the last International Congress of Mathematicians, Daniel Gorenstein described the advanced state of our efforts to classify the finite simple groups. A number of problems remained, including some component problems and work on groups of characteristic two type, the characterization of the Ree groups and the construction of two of the sporadic simple groups. By February, 1981, these points had been settled and so a proof of the classification of finite simple groups was completed! While, on one hand, things seemed to be winding down, a remarkable and unexpected link between the sporadic groups and modular forms was discovered and promises to stimulate a great deal of research.

I wish to talk mainly about the sporadic simple groups and some recent developments in the theory of finite simple groups. Since Gorenstein's enthusiastic account of the classification has appeared [25], I shall say little about that. I shall present some information about the finite simple groups and make a few comments.

2. The classification of finite simple groups. The lengthy proof of the classification of finite simple groups exists in a combination of published articles and preprints. The finite group theory community is confident in the program. Though we expect to find errors as the proofs are reexamined, we feel that the well-established techniques will suffice to straighten out the problems which arise. I refer to Gorenstein's book [25] and Feit's review [17] for two discussions of this point. My thoughts on this are generally compatible with theirs. However, I am not sure that the list of simple groups will remain unchanged by future examinations of the arguments. The proof is incredibly long and delicate.

The following table is taken from [25]. For a group of Lie type, the order of the "simply connected" version is given. The "adjoint version" (simple in all but a few cases) is obtained via factoring by a central subgroup of order d, given in the right column.

The finite simple groups and their orders

Group G	Order of G	d
	Groups of Lie type	
$A_n(q)$	$q^{n(n+1)/2}\prod_{i=1}^{n}(q^{i+1}-1)$	$(n+1, q-1)$
$B_n(q),\ n>1$	$q^{n^2}\prod_{i=1}^{n}(q^{2i}-1)$	$(2, q-1)$
$C_n(q),\ n>2$	$q^{n^2}\prod_{i=1}^{n}(q^{2i}-1)$	$(2, q-1)$
$D_n(q),\ n>3$	$q^{n(n-1)}(q^n-1)\prod_{i=1}^{n-1}(q^{2i}-1)$	$(4, q^n-1)$
$G_2(q)$	$q^6(q^6-1)(q^2-1)$	1
$F_4(q)$	$q^{24}(q^{12}-1)(q^8-1)(q^6-1)(q^2-1)$	1
$E_6(q)$	$q^{36}(q^{12}-1)(q^9-1)(q^8-1)(q^6-1)(q^5-1)(q^2-1)$	$(3, q-1)$
$E_7(q)$	$q^{63}(q^{18}-1)(q^{14}-1)(q^{12}-1)(q^{10}-1)(q^8-1)(q^6-1)(q^2-1)$	$(2, q-1)$
$E_8(q)$	$q^{120}(q^{30}-1)(q^{24}-1)(q^{20}-1)(q^{18}-1)(q^{14}-1)(q^{12}-1)(q^8-1)(q^2-1)$	1
${}^2A_n(q),\ n>1$	$q^{n(n-1)/2}\prod_{i=1}^{n}(q^{i-1}-(-1)^{i-1})$	$(n+1, q+1)$
${}^2B_2(q),\ q=2^{2m-1}$	$q^2(q^2+1)(q-1)$	1
${}^2D_n(q),\ n>3$	$q^{n(n-1)}(q^n+1)\prod_{i=1}^{n-1}(q^{2i}-1)$	$(4, q^n+1)$
${}^3D_4(q)$	$q^{12}(q^8+q^4+1)(q^6-1)(q^2-1)$	1
${}^2G_2(q),\ q=3^{2m-1}$	$q^3(q^3+1)(q-1)$	1
${}^2F_4(q),\ q=2^{2m-1}$	$q^{12}(q^6+1)(q^4-1)(q^3+1)(q-1)$	1
${}^2E_6(q)$	$q^{36}(q^{12}-1)(q^9+1)(q^8-1)(q^6-1)(q^5+1)(q^2-1)$	$(3, q+1)$
	Alternating groups	
$A_n,\ n\geqslant 5$	$\frac{1}{2}(n!)$	
	Sporadic groups	
M_{11}	$7920 = 2^4\cdot 3^2\cdot 5\cdot 11$	
M_{12}	$95040 = 2^6\cdot 3^3\cdot 5\cdot 11$	
M_{22}	$443520 = 2^7\cdot 3^2\cdot 5\cdot 7\cdot 11$	
M_{23}	$10200960 = 2^7\cdot 3^2\cdot 5\cdot 7\cdot 11\cdot 23$	

M_{24}	$244823040 = 2^{10}\cdot 3^3\cdot 5\cdot 7\cdot 11\cdot 23$
J_1	$175560 = 2^3\cdot 3\cdot 5\cdot 7\cdot 11\cdot 19$
J_2	$2^7\cdot 3^3\cdot 5^2\cdot 7$
J_3	$2^7\cdot 3^5\cdot 5\cdot 17\cdot 19$
J_4	$2^{21}\cdot 3^3\cdot 5\cdot 7\cdot 11^3\cdot 23\cdot 29\cdot 31\cdot 37\cdot 43$
HS	$2^9\cdot 3^2\cdot 5^3\cdot 7\cdot 11$
Mc	$2^7\cdot 3^6\cdot 5^3\cdot 7\cdot 11$
Suz	$2^{13}\cdot 3^7\cdot 5^2\cdot 7\cdot 11\cdot 13$
Ru	$2^{14}\cdot 3^3\cdot 5^3\cdot 7\cdot 13\cdot 29$
He	$2^{10}\cdot 3^3\cdot 5^2\cdot 7^3\cdot 17$
Ly	$2^8\cdot 3^7\cdot 5^6\cdot 7\cdot 11\cdot 31\cdot 37\cdot 67$
ON	$2^9\cdot 3^4\cdot 5\cdot 7^3\cdot 11\cdot 19\cdot 31$
.1	$2^{21}\cdot 3^9\cdot 5^4\cdot 7^2\cdot 11\cdot 13\cdot 23$
.2	$2^{18}\cdot 3^6\cdot 5^3\cdot 7\cdot 11\cdot 23$
.3	$2^{10}\cdot 3^7\cdot 5^3\cdot 7\cdot 11\cdot 23$
$M(22)$	$2^{17}\cdot 3^9\cdot 5^2\cdot 7\cdot 11\cdot 13$
$M(23)$	$2^{18}\cdot 3^{13}\cdot 5^2\cdot 7\cdot 11\cdot 13\cdot 17\cdot 23$
$M(24)'$	$2^{21}\cdot 3^{16}\cdot 5^2\cdot 7^3\cdot 11\cdot 13\cdot 17\cdot 23\cdot 29$
F_5	$2^{15}\cdot 3^{10}\cdot 5^3\cdot 7^2\cdot 13\cdot 19\cdot 31$
F_3	$2^{14}\cdot 3^6\cdot 5^6\cdot 7\cdot 11\cdot 19$
F_2	$2^{41}\cdot 3^{13}\cdot 5^6\cdot 7^2\cdot 11\cdot 13\cdot 17\cdot 19\cdot 23\cdot 31\cdot 41\cdot 47$
F_1	$2^{46}\cdot 3^{20}\cdot 5^9\cdot 7^6\cdot 11^2\cdot 13^3\cdot 17\cdot 19\cdot 23\cdot 29\cdot 31\cdot 41\cdot 47\cdot 59\cdot 71$

3. Sporadic simple groups. The term "sporadic simple group" comes from a remark of Burnside in one of his famous books on finite group theory [4], page 504. He meant, by the term, a finite simple group not in one of the naturally defined infinite families. He knew of five sporadic groups, the Mathieu groups [50]. Since Burnside's day, further infinite families of simple groups were discovered, but none which contained the five groups already declared sporadic. See the papers of Chevalley [5], Ree [58], [59], Steinberg [62] and Suzuki [63] (but note that Dickson knew of the groups of type G_2 [14] and E_6 [15], defined over any field).

The picture of the known simple groups changed radically in the 1960's. Janko's paper [38] describes a finite simple group, usually denoted J_1, of order $175560 = 2^3\cdot 3\cdot 5\cdot 7\cdot 11\cdot 19$. It was encountered as a solution to the problem of determining simple groups which have abelian Sylow 2-groups and which contain an involution whose centralizer is isomorphic to $Z_2\times \mathrm{PSL}(2, q)$, for odd $q > 3$; the groups of Ree type occur for the case $q = 3^{2n+1}$, $n > 1$ [41] and J_1 occurs for $q = 5$ [38]. The discovery of this group marked the beginning of a decade of discoveries of further sporadic groups. Evidence for the twenty sixth and last group was, appropriately enough, announced by Janko in the spring of 1975 [39]. Naturally, the possibilities of further simple groups were most exciting. Many candidates for centralizers of involutions in simple groups were subjected to the rigors of "local analysis". By far, most of the candidates led to contradictions, but a very few special ones did not. Other ways sporadic groups were found were in studies of graphs and their automorphisms, especially "rank 3 graphs" [35], [36], groups generated by classes of ω-transpositions [18], [19], and automorphisms of lattices, mainly the Leech lattice and its sublattices [7], [8]. As with the centralizer of involution method, many possibilities in these areas were tried as sources for new groups. Suffice it to say that the successes were beautiful but small in number.

II. Recent developments

4. The Ree group problem. I have already mentioned the centralizer of involution problem associated to the simple groups of Ree [40]. The early work on this problem led to a group order and a lot of internal information, but not to the uniqueness of its isomorphism type. Thompson worked quite hard on this, eventually producing a set of algebraic equations over $\boldsymbol{F}_q$, $q = 3^{2n+1}$ [68]. He had already shown that the isomorphism type of the group depended only on an automorphism σ of this field, and he felt that these equations should determine σ, i.e. should force $\sigma^2 = 3$. This problem resembles those of Suzuki [63] and O'Nan [56], but is vastly more difficult owing to the fact that the nilpotence class of the relevant Sylow group is 3 (it is 2 in the other cases). The final step in settling the problem was taken by Bombieri [1], who applied classical elimination theory to study the equations and force the correct answer. The small fields were troublesome, and had to be handled separately by machine. Andrew Odlyzko and David Hunt independently checked the relevant cases. Thompson's work on this problem stretched over a decade, a real act of perseverance! The work of both Thompson and Bombieri were displays of considerable insight and technical power!

5. Connections with modular forms. This was very dramatic. John McKay's observation that $196884 = 1+196883$ started it all. In the mid 1970's, we knew that the putative simple group, F_1 (see Section 6), had the following property: a nonprincipal irreducible matrix representation had degree at least $196883 = 47\cdot 59\cdot 71$. I conjectured that this number was indeed a degree [30] (see also [9]); it is, but more on this later. McKay noticed that the elliptic modular function $j(\tau)$ had a very interesting coefficient in its series expansion with respect to the variable $q = e^{2\pi i\tau}$:

$$j(\tau) = q^{-1} + 744 + 196884q + 21493760q^2 + \dots$$

In late 1978, Thompson [67] suggested that the coefficients may be simple positive integral linear combinations of irreducible degrees for F_1 and that we ought to replace them by one of the representations which this process suggests. Conway and Norton [9] studied the list of conjugacy classes of F_1 and a list of function fields of genus 0 and found that there was, roughly speaking, a bijection between the rational conjugacy classes of F_1 and a list of these function fields. The correspondence may be expressed this way. There are class functions H_n, $n \geqslant 1$, such that

$$g \mapsto T_g := q^{-1} + \sum_{n=1}^{\infty} H_n(g) q^n, \qquad \text{for } g \in F_1,$$

gives the above correspondence. The object

$$T := q^{-1} + \sum_{n=1}^{\infty} H_n q^n$$

with coefficients H_n in the ring of class functions, is called the *Thompson series*. By definition, its evaluation on a group element gives, up to the addition of a constant, a Hauptmodul, i.e., a normalized generator of one of the aforementioned function fields. Thompson's suggestion was converted to the following conjecture: the H_n are characters of representations of F_1. Assuming the validity of the character table of F_1 [21], A. O. Atkin, Paul Fong and Steve Smith verified the conjecture in 1979, using arguments which combined analysis of coefficients of the Hauptmoduln, group theoretic methods and computer programming.

So many questions are raised by this contact between finite groups and modular forms! It is widely felt that something very deep is going on here. An explanation of why and how the simple groups involved in F_1 are related to the genus 0 function fields will be a challenge to mathematicians for years to come.

Another "numerology" result linking sporadic groups with number theory was obtained by Geoffrey Mason [49]. He finds that the conjugacy classes of M_{24} are more-or-less in bijection with a family of cusp forms which have an Euler product expansion.

Recent developments in the theory of infinite dimensional Lie algebras (see [43] and [47] for example) have included explicitly described graded spaces whose generating functions are modular forms. When the story about the Thompson series began to get around, several mathematicians defined spaces which had essentially $q \cdot j(\tau)$ as the generating function [23], [42], [47]. Further spaces have been defined. It seems likely that an appropriate graded structure would help us to understand the relationship between sporadic groups and modular forms. In October, 1983, Frenkel, Lepowsky and Meurman announced the construction of a graded F_1-module with $q\cdot(j(\tau)-744)$ as its generating function [71].

6. Construction of the largest sporadic simple group. In November, 1973, Bernd Fischer and I independently produced evidence for the existence of a finite simple group of order

$2^{46}\cdot 3^{20}\cdot 5^{9}\cdot 7^{6}\cdot 11^{2}\cdot 13^{3}\cdot 17\cdot 19\cdot 23\cdot 29\cdot 31\cdot 41\cdot 47\cdot 59\cdot 71 =$
808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000.

We shall refer to any finite simple group of this order as a group of type F_1; see [27], [25]. In early 1980, I constructed a finite simple group of type F_1. There is a uniqueness result, due to Thompson [66] and Norton [53]. By combining it with results from the classification theory, we may assert that a group of type F_1 is unique up to isomorphism.

I like the optimistic name "Friendly Giant" for this simple group. However, the name "monster" has been around for a while and continues to grow more popular. Accordingly, I have decided to use it.

In the years following the discovery of F_1, there were some doubts about the possibility of a construction [26] because of the sort of constructions made for sporadic groups in the past. For example, the biggest subgroup of F_1 we know of is $2\cdot F_2$, a perfect central extension of the Fischer group F_2. The index is about 10^{20}. So, if one were to describe F_1 as a group of permutations, it would be necessary to define permutations on about 10^{20} symbols and be able to compute useful things with them. This is probably beyond what one can do with computing machines at the present time.

My method was to construct an algebra, B, of dimension 196884 over Q, which is commutative and nonassociative, then describe a group G of algebra automorphisms. Techniques from the classification of finite simple groups were invoked to prove finiteness of G and identify G as a finite simple group of the correct order. Let B_0 be the 196883 dimensional module with unique (up to scalar multiplication) algebra structure invariant under G; it is derived from B by projecting products onto B_0. J. Tits has recently proved that $\mathrm{Aut}(C\otimes B_0)$ is finite and, by invoking arguments from [28], concludes that $\mathrm{Aut}(C\otimes B_0)=G$.

Two basic ingredients to my construction should be singled out: extraspecial 2-groups [24], [37] and the Leech lattice [45], [46]. A finite p-group Q is *extraspecial* if and only if $Q'=Z(Q)\cong Z_p$ ($'$ denotes the commutator subgroup). Such a group has order p^{1+2n}, for some integer $n\geqslant 1$, p linear characters and $p-1$ nonlinear characters, each of degree p and forming a complete algebraically conjugate set. The *Leech lattice* L is a remarkable rank 24 lattice. It is the only one of the 24 rank 24 even unimodular lattices [52], [69] with no vector of squared length 2; the minimal vectors have squared length 4. Though no reflection on $\boldsymbol{R}\otimes L$ preserves L, $\mathrm{Aut}(L)$ is a remarkable finite group, called $\cdot 0$ ("dot zero"). Modulo its center $\{\pm 1\}$, it is the simple group $\cdot 1$ of Conway, of order $2^{21}\cdot 3^{9}\cdot 5^{4}\cdot 7^{2}\cdot 11\cdot 13\cdot 23$ [7], [8]. Essential to understanding the Leech lattice and the Conway groups are the *Mathieu groups* [50], especially M_{27} and its *Steiner system* $S = S(5, 8, 24)$. The simple group M_{24} has order $2^{10}\cdot 3^{3}\cdot 5\cdot 7\cdot 11\cdot 23$ and operates as a 5-transitive permutation group on Ω, a set of 24 letters. It may be defined as the set of permutations on Ω which preserve S, a collection of 8 element sets, called *octads*, which have the following property: exactly one octad contains any given five points. Such a system exists and is unique up to isomorphism [70]. Easily, we count $\binom{24}{5}\Big/\binom{8}{5} = 759$ octads. In the power set on Ω, a vector space under symmetric difference, the octads span a 12-dimensional space, called the *Golay code* or the *$\mathscr{C}$-sets*. The cardinalities of sets in this space are 0, 8, 12, 16 and 24 and there are 1, 759, 2576, 759, and 1 such sets, respectively.

I shall give a brief description of B. In F_1, there is an involution z whose centralizer C has shape $2^{1+24}\cdot(\cdot 1)$. There is no problem giving an abstract construction of C and understanding its representation theory well enough to write down a module for C: $B = U\oplus V\oplus W$, where $U\oplus V$ is the 1 eigenspace of z and W the -1 eigenspace, and where $U = S^2(H)$, where $H = \boldsymbol{Q}\otimes L$ and C operates on U via the natural quotient $C\to\cdot 1$;

V has basis corresponding to all pairs $\{X, -X\}$, where X is a short vector in L (i.e., $(X, X) = 4$) and C operates monomially with respect to this basis;

$W = H\otimes T$, where the covering group $\hat{C}$ of C operates on the two factors, inducing $\cdot 0$ on H and a linear group on T (dimension 2^{12}) which extends the natural faithful action of an extraspecial group 2^{1+24} on T and, like C, is an extension $2^{1+24}\cdot(\cdot 1)$.

We have $\dim U = 300$, $\dim V = 98280$ and $\dim W = 98304$. Both V and W are irreducible, while U has constituents of dimension 1 and 299; call them $\boldsymbol{Qd}$ and U_0. I described all C-maps $X\times Y\to Z$, where X, Y and Z range over these four irreducibles, and I chose a system of these maps to define an algebra structure on B. My choices were motivated by a wish to make the algebra structure invariant under some linear transformation σ which satisfies $\sigma^{-1}z\sigma \neq z$. The definition of a suitable σ was the biggest single problem here.

The group G (see three paragraphs before) is, by definition, the group generated by C and σ.

The Leech lattice L may be defined as the lattice in 24-space spanned by its minimal vectors, L_2, which may be partitioned into sets L_2^4, L_2^2 and L_2^3 as follows (I omit a scale factor of $1/\sqrt{8}$):

L_2^4 = all vectors of shape $(4^2 0^{22})$, $\binom{24}{2}4 = 1104$ in number;

L_2^2 = all vectors of shape $(2^8 0^{16})$ with support an octad and an even number of minus signs, $759 \cdot 128 = 97152$ in number;

L_2^3 = all vectors obtained from $(-3, 1, 1, \ldots, 1)$ by changing signs at all $\mathscr{C}$-sets (the -3 is allowed at all coordinate positions), $24 \cdot 4096 = 98304$ in number.

The total is $196560 = |L_2|$. The decomposition of B was refined as follows:

$U = B_{24} \oplus B_{276}$; the subscript indicates the dimension and B_{24} is the span of all x_i^2, $i \in \Omega$, and B_{276} is the span of all $x_i x_j$, $i \neq j$ in Ω;

$V = B_4 \oplus B_2 \oplus B_3$, where B_k is the span of all the 1-dimensional spaces indexed by $\{X, -X\}$, $X \in L_2^k$; furthermore, $B_4 = B_4^+ \oplus B_4^-$, where B_4^+, B_4^- is the span of all $v+v'$, $v-v'$, respectively, where v, v' range over all pairs of the canonical basis vectors corresponding to pairs $\{X, -X\}$, $\{Y, -Y\}$ with $X = (4, 4, 0, \ldots, 0)$, $Y = (4, -4, 0, \ldots, 0)$, both supported at the same 2-set.

$W = H \otimes (T(1) \oplus T(2))$, where the right hand factor is a decomposition of T into 1 and -1 eigenspaces for a noncentral involution of the extraspecial group 2^{1+24}.

An explicit description of the "extra automorphism" σ was made using the bases associated to the above refinement. I must refer to [27], [28] for details, but can suggest its definition:

$$\sigma|_{B_{24}} = 1, \quad \sigma|_{B_4^+} = 1, \quad \sigma\colon B_{276} \leftrightarrow B_4^-;$$

σ preserves all the 64-dimensional spaces in B_2 which correspond to octads;

$$\sigma = \pm 1\text{'s on a canonical basis for } H \otimes T(2);$$

$$\sigma\colon B_3 \xleftrightarrow{\pm 1} H \otimes T(1).$$

It is possible to modify the definition of B given in [27], [28] to keep G as algebra automorphisms and make $\boldsymbol{d}$ act as the identity (this may be done in more than one way). Arne Meurman pointed out that U may be made to look like the Jordan algebra of 24 by 24 symmetric matrices in a suitably modified B. However, B is not a Jordan algebra. A linear homogeneous identity satisfied by B in commuting, nonassociating variables has not been found; one must exist and have degree at least 6 [72]. Other descriptions and analyses of B have been made by J. Tits and by I. Frenkel, J. Lepowsky and A. Meurman. One would like to know things like the linear identities satisfied by B, properties of idempotents, etc. However, "classical" questions seem unnatural (so far) because of the way B came to exist, as a "minimal object" needed to display a group.

Although B is not a classical nonassociative algebra, one can point to a few similarities. For example, if $V = \oplus V_i$ is a decomposition using the above basis, we have $\dim V_i = 1$ and $U \cdot V_i = V_i$, so that the V_i look like "root spaces" for U. There are several reasons for thinking of W as a "spin module".

It is amusing to note that B (the version which makes U a Jordan algebra) contains a linearly independent set S of 48 vectors with the property that if $R \subseteq S$, the linear span of U and R is a simple subalgebra of dimension $300 + |R|$. This strikes me as rather unclassical.

7. Consequences. From the existence of F_1, we may derive the existence of other sporadic simple groups as sections (quotients of subgroups). In F_1, one finds twenty of the twenty six sporadics as sections (I call these the "Happy Family"). Five sporadics are not involved in F_1 (these are the "Pariahs"). Whether J_1 is a section in F_1 is not settled.

Since we used C heavily in the construction and $C/O_2(C) = \cdot 1$, we cannot claim new existence proofs for the groups involved in C. What we get are new (and almost trivial) existence proofs for the Fischer groups F_{22}, F_{23} and F_{24}; also, the group F_2 of Fischer, F_3 of Thompson, F_5 of Harada and Norton and F_7 of Held. It is a simple matter to name elements $x \in C$ so that $C_G(x)/\langle x \rangle$ is one of F'_{24}, F_2, F_3, F_5 and F_7; one may see F_{22} and F_{23} as sections within F'_{24}. Constructions of these last four groups had been done one by one, with computers. The first three groups had been constructed by Fischer in his beautiful and important work on 3-transposition groups [18], [19]. My existence proof is in a completely different style and it is a great pleasure for me to see the two points of view make contact here. The groups F_7 and F_2 had been constructed as permutation groups on the cosets of subgroups isomorphic to $\mathrm{Sp}(4, 4) \cdot 2$ and $2^2 \cdot E_6(2) \cdot 2$, respectively, while F_3 and F_5 had been constructed as linear groups in dimensions 248 and 133, respectively (despite the numbers, these representations do not embed these groups in $E_8(C)$ and $E_7(C)$, respectively; however, by reducing the representation of F_3 modulo 3, we get an embedding in $E_8(3)$).

Uniqueness results are available for all these groups, but none resulted from my work. It is quite reasonable to expect uniqueness results from a study of various low-dimensional representations of these groups, say, in the style of [22] or [66].

The existence of F_1 also implies the nonvanishing of certain cohomology groups. See item 5 in the last section for some remarks on this.

8. Variations and generalizations. This algebra provides examples of many finite groups acting on (non) associative algebras. If S is a subgroup of F and A is an S-submodule of B, we get an algebra structure on A by projecting products onto A and S operates on A as algebra automorphisms.

Other examples of finite groups (essentially simple) as automorphisms of nonassociative algebras are available, most notably the groups of Lie type and their Lie algebras. Besides these, there are families of commutative nonassociative algebras derived from permutation groups [54], [61]. We sketch the example associated to a 3-transposition group. Let D be a class of 3-transpositions generating the finite group G and let $\{x_d\colon d \in D\}$ be the usual orthonormal basis for the permutation module. Define a product $*$ on the module by declaring $x_d * x_e$ to be x_f if $\{d, e, f\}$ are the involutions of a dihedral group of order 6 and 0 otherwise. One can then take a nontrivial constituent A of the module and project products onto it, thus making A an algebra; usually $A \cdot A \neq 0$. Finally, I mention the recent construction by Frohardt of a 170 dimensional algebra and a group of automorphisms isomorphic to J_3 [22]. The group is small, compared to F_1, but is relatively hard to construct in this manner. In the construction of F_1, one had to choose 6 parameters to give a C-map which makes B have the right algebra structure. The corresponding number for this construction of J_3 is 51. Other work along this line is in progress. Infinite dimensional examples are being considered too. I expect a theory to take shape in time.

9. The construction of Janko's fourth group. The group J_4, conjectured to exist by Janko in 1975 [39] and having order $2^{21}\cdot 3^3\cdot 5\cdot 7\cdot 11^3\cdot 23\cdot 29\cdot 31\cdot \cdot 37\cdot 43$, was constructed by Simon Norton with the assistance of R. A. Parker, J. H. Conway, D. J. Benson and J. G. Thackray. Details have not been circulated although an outline has appeared [55]. Let G be a finite simple group satisfying Janko's hypotheses [39]. The smallest degree of a nonprincipal complex irreducible representation is 1333. Thompson suggested that G might have a representation of degree 112 over F_2. In any case, G must contain subgroups of shape $2^{1+12}3\cdot M_{22}\cdot 2$ and $2^{11}M_{24}$.

Matrices in GL(112, 2) generating two such subgroups were defined. Pure group theory and computer work determined that these two subgroups generate a group satisfying Janko's hypotheses [39]. Many properties of the Mathieu groups were used throughout the analysis. A PSU(3, 11) subgroup of G was also important. A uniqueness result (that G is unique up to isomorphism) was obtained.

III. The present moment

At this time, the most dramatic challenge is to explain the connection between F_1 (the "Friendly Giant", or the "Monster", as you prefer) and the modular forms associated to groups of genus zero. I have already discussed this and related matters. Many other questions and observations about the sporadic simple groups tantalize us. I give a brief sample, chosen mainly to exhibit variety. (I do not want to give the impression that I am an authority on the areas I mention below.) A more thorough account of how we currently see the relations of finite simple groups to other areas of mathematics may be found in the proceedings of last summer's group theory conference in Montreal [51].

1. What sort of interesting connections are there between invariant forms of degree 3 or more and internal properties of simple groups? For instance, neat formulas giving cubic forms may be associated to 3-transposition groups [54], [61]. In the notation of the previous section, a cubic form on A is defined by $(y_d, y_e, y_f) = (y_d * y_e, y_f)$, where $(\ ,\ )$ is the usual bilinear form and the y's are the projections of the x's onto A; note that the right side really is symmetric in d, e, f. Also, note that the algebras used to construct F_2 and J_3 have invariant cubic forms, due to the fact that the bilinear forms are associative. It is not known whether the forms are nonsingular.

2. We have had a classification of finite irreducible real and complex reflection groups for some time, and their connections with invariant theory are well-known. A nonsolvable composition factor of such a group is an alternating group or is on a short list of classical groups over small finite fields. Recently, A. Cohen [6] classified quaternionic reflection groups. A few more composition factors occur, including the sporadic group HJ of Hall and Janko [32], [40]. The finding of HJ in this "classical" setting seems to challenge the meaning of "sporadic"!

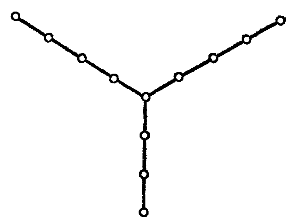
3. John McKay furnishes the following observation. Take an involution from the conjugacy class $2A$ in F_1. The centralizer $C(x)$ acts by conjugation on this class. Let $x_1, x_2, \ldots$ represent the orbits distinct from $\{x\}$ and define

the integer $n_i = |xx_i|$. Then it seems (though I am not sure that it has been rigorously checked) that the n_i may be arranged as follows:

$$\begin{array}{ccccccc} 2 & 4 & 6 & 5 & 4 & 3 & 2 \\ & & 3 & & & & \end{array}$$

This displays the coefficients of the highest root in E_8!

4. Bernd Fischer has studied a diagram



with the property that the corresponding reflection group, W, maps onto F_1 in such a way that the reflections corresponding to the nodes map to involutions in the class $2A$. Let K be the kernel; K is not a "congruence subgroup" [31].

Diagrams of the above general shape (three branches, single bonds only) are familiar objects in algebra, e.g. [2], and in algebraic geometry, particularly in the theory of singularities [10], [11], [16], [48]. It is natural to wonder if the sequence $1 \to K \to W \to F_1 \to 1$ is connected to some interesting algebraic geometry.

5. McLaughlin's observation. It is curious that sporadic groups are associated to so many examples of nonvanishing cohomology. The existence of F_1 implies the existence of a number of nonsplit extensions, hence nonvanishing degree 2 cohomology. For example, $3^8 \cdot O^-(8,3)$ and $2 \cdot F_2$ occur as subgroups (since $H^2(O^\varepsilon(n,3), F_3^n)$ is almost always zero [44], the nonsplit extension $3^8 \cdot O^-(8,3)$ is regarded as "exceptional"). There are lots of examples involving trivial and nontrivial modules. Centralizers of involutions are a central concept in the classification of simple groups, and one finds in sporadic groups exceptional coverings of groups of Lie type and central extensions of sporadic groups as centralizers of involutions. Here is an example involving nontrivial extensions. Define $d_n = \dim H^2(\mathrm{GL}(n,2), F_2^n)$. Work of Dempwolff [12], [13] and others shows that $d_n = 1$ for $n = 3, 4, 5$ and is 0 otherwise. Let E_n be the relevant nonsplit extension for $n = 3, 4, 5$. Then E_3, E_4 and E_5 occur as maximal

2-locals in $G_2(q)$ for q odd, $\cdot 3$ and F_5, respectively. The extension E_5 is usually called the *Dempwolff extension*. It is connected with a nonstandard decomposition of the Lie algebra E_8 [64] and has index 2^{10} in a maximal finite subgroup of the group $E_8(C)$ [29], [64].

6. Is there a structure theory of sporadic simple groups which would allow one to make efficient and uniform arguments (as with the groups of Lie type) and would provide new insights? Work on automorphism groups of algebras and on finite geometries (e.g., [3], [60]) both touch (nearly all) the sporadic groups and may eventually give us something along this line. At the moment, we are not even close to formulating useful axioms.

IV. Concluding remarks

I have concentrated on aspects of finite group theory of special interest to me and will not try to detail all the current work. The classification is being revised and reexamined. Improvements on the odd order theorem, the dihedral and semidihedral papers and the sectional 2-rank 4 theorem should be mentioned. Representation theory is quite active. There have been results on Galois groups and permutation groups.

Though finite group theory has been permanently changed by the classification of finite simple groups, it continues to show strong character. The future will be rich in excitement and mysteries, as was the past.

References

[1] Bombieri Enrico, Thompson's Problem ($\sigma^2 = 3$), *Invent. Math.* **58** (1980), pp. 77–100.
[2] Bourbaki Nickolai, *Groupes et Algèbres de Lie*, Chapitres 4, 5 and 6, Hermann, Paris, 1968.
[3] Buekenhout Francis, Diagrams for Geometries and Groups, *J. Combinatorial Theory* **27** (1979), pp. 121–151.
[4] Burnside William, *Theory of Groups*, 2nd edition, Cambridge University Press, 1911 (Dover, New York, 1955).
[5] Chevalley Claude, Sur certains groupes simples, *Tôhoku Math. J.* **7** (1955), pp. 14–66.
[6] Cohen Arjeh, Finite Quaternionic Reflection Groups, *J. Algebra* **64** (1980).
[7] Conway John, A Group of Order 8, 315, 553, 613, 086, 720, 000, *Bull. London Math. Soc.* **1** (1969), pp. 79–88.
[8] Conway John, Three Lectures on Exceptional Groups, In: G. Higman and M. Powell (eds.), *Finite Simple Groups, Oxford, 1969*, Academic Press, London, 1971.

[9] Conway John and Norton Simon, Monstrous Moonshine, *Bull. London Math. Soc.* **11** (1979), pp. 308–339.
[10] Coble A., *Algebraic Geometry and Theta Functions*, A. M. S. Colloq. Publications, X, Providence, 1929 (second edition, 1961).
[11] Coble A., Theta Modular Groups Determined by Point Sets, *Amer. J. Math.* **114** (1918), pp. 267–322.
[12] Dempwolff Ulrich, On the Extensions of an Elementary Group of Order 2^5 by GL(5,2), *Rend. Sem. Mat. Univ. Padova* **48** (1972), pp. 359–364.
[13] Dempwolff Ulrich, On the Second Cohomology of GL(n, 2), *J. Austral. Math. Soc.* **16** (1973), pp. 207–209.
[14] Dickson Leonard Eugene, A New System of Simple Groups, *Math. Ann.* **60** (1905), pp. 137–150.
[15] Dickson Leonard Eugene, A Class of Groups in an Arbitrary Realm Connected with the Configuration of the 27 Lines on a Cubic Surface. Second paper, *Quarterly Journal of Pure and Applied Mathematics* **39** (1908), pp. 121–148.
[16] Dolgachev Igor, Weyl Groups and Cremona Transformations, *Proc. Summer Institute on Singularities, Arcata, 1981* (to appear).
[17] Feit Walter, Review of "Finite Simple Groups", by Daniel Gorenstein, *Bull. Amer. Math. Soc.* **8** (1983), pp. 120–124.
[18] Fischer Bernd, *Finite Groups Generated by 3-Transpositions*, Univ. of Warwick preprint, 1969.
[19] Fischer Bernd, Finite Groups Generated by 3-Transpositions, *Invent. Math.* **13** (1971), pp. 232–246.
[20] Fischer Bernd, privately circulated note on F_1, 1973.
[21] Fischer B., Livingstone, D. and Thorne, *The Character Table of* F_1.
[22] Frohardt Daniel, A Trilinear Form for J_3, to appear in *J. Algebra.*
[23] Garland Howard, lectures, 1980.
[24] Gorenstein Daniel, *Finite Groups*, Harper and Row, New York, 1968; 2nd ed., Chelsea, New York, 1980.
[25] Gorenstein Daniel, *Finite Simple Groups*, Plenum Press, New York, 1981.
[26] Gorenstein Daniel, *Proceedings of the International Congress of Mathematicians, Helsinki 1978*, Academia Scientarum Fennica, 1980.
[27] Griess Robert L., Jr., A Construction of F_1 as Automorphisms of a 196883 Dimensional Algebra, *Proc. Nat. Acad. Sci. U.S.A.* **78** (1981), pp. 689–691.
[28] Griess Robert L., Jr., The Friendly Giant, *Invent. Math.* **69** (1982), pp. 1–102.
[29] Griess Robert L., Jr., On a Subgroup of Order $2^{15}|\mathrm{GL}(5,2)|$ in $E_8(\boldsymbol{C})$, *J. Algebra* **40** (1976), pp. 271–279.
[30] Griess Robert L., Jr., The Structure of the "Monster" Simple Group. In: W. Scott and F. Gross (eds.), *Proceedings of the Conference on Finite Groups*, Academic Press, New York, 1976, pp. 113–118.
[31] Griess Robert L., Jr., Quotients of Infinite Reflection Groups, to appear in *Math. Ann.*
[32] Hall Marshall, Jr. and Wales David, The Simple Group of Order 604, 800, *J. Algebra* **9** (1968), pp. 417–450.
[33] Harada Koichiro, On the Simple Group F of Order $2^{14}\cdot 3^6\cdot 5^6\cdot 7\cdot 11\cdot 19$. In W. Scott and F. Gross (eds.), *Proceedings of a Conference on Finite Groups*, Academic Press, New York, 1976, pp. 119–195.

[34] Held Dieter, The Simple Groups Related to M_{24}, *J. Algebra* **13** (1969), pp. 253–296.
[35] Higman D. G., A Survey of Some Questions and Results about Rank 3 Permutation Groups. In: *Actes du Congrès Intern. Math.*, Tome 1, Gauthier-Villars, Paris, 1971, pp. 361–365.
[36] Higman D. G., Finite Permutation Groups of Rank 3, *Math. Z.* **86** (1964), pp. 145–156.
[37] Huppert Bertram, *Endliche Gruppen*, I, Springer-Verlag, Berlin and New York, 1967.
[38] Janko Zvonimir, A New Finite Simple Group with Abelian 2-Sylow Subgroup and its Characterization, *J. Algebra* **3** (1966), pp. 147–186.
[39] Janko Zvonimir, A New Finite Simple Group of Order 86, 755, 571, 046, 077, 562, 880 which Posesses M_{24} and the Full Cover of M_{22} as Subgroups, *J. Algebra* **42** (1976), pp. 564–596.
[40] Janko Zvonimir, Some New Simple Groups of Finite Order I, *First Naz. Alta Math. Symposia Math.* **1** (1968), pp. 25–65.
[41] Janko Z. and Thompson J. G., On a Class of Finite Simple Groups of Ree, *J. Algebra* **4** (1966), pp. 274–292.
[42] Kac Victor G., An Elucidation of "Infinite Dimensional Algebras ... and the Very Strange Formula," $E_8^{(1)}$ and the Cube Root of the Modular Invariant j, *Advances in Math.* **3** (1980), pp. 264–273.
[43] Kac Victor G., Highest Weight Representations of Infinite Dimensional Lie Algebras. In: *Proceedings of the International Congress of Mathematicians, Helsinki 1978*, Academia Scientarum Fennica, 1980.
[44] Kusefoglu Ayse, Cohomology of Finite Orthogonal Groups, I, *J. Algebra* **56**, pp. 207–220; *J. Algebra* **67** (1980), pp. 88–109.
[45] Leech John, Notes on Sphere Packings, *Canad. J. Math.* **19** (1967), pp. 251–267.
[46] Leech John, Some Sphere Packings in Higher Space, *Canad. J. Math.* **16** (1964), pp. 657–682.
[47] Lepowsky James, Euclidean Lie Algebras and the Modular Function j. In: *The Santa Cruz Conference on Finite Groups*, Proc. Symp. Pure Math. 37, American Mathematical Society, Providence, 1980.
[48] Looijenga E., Rational Surfaces with an Anticanonical Cycle, *Ann. of Math.* **114** (1981), pp. 267–322.
[49] Mason Geoffrey, M_{24} and Certain Automorphic Forms, to appear in the *Proceedings of the Montreal Group Theory Conference*, 1982.
[50] Mathieu Emil, Memoire sur la nombre de valeurs que peut acquirer une fonction quand on y permut ses variables de toutes les manieres possibles, *Liouville's Journal* **5** (1860), pp. 9–42; Memoire sur l'etude des fonctions de plusieurs quantites, sur la maniere de les formes et sur les substitutions qui les laissent invariables, *Liouville's Journal* **6** (1861), pp. 241–323; Sur la fonction cinq fois transitives des 24 quantites, *Liouville's Journal* **18** (1873), pp. 25–46.
[51] McKay John (ed.), *Proceedings of the 1982 Conference on Finite Groups, held in Montreal*; to appear.
[52] Niemeyer H. V., *Definite Quadratische Formen der Diskriminante* 1 *und Dimension* 24, Doctoral Dissertation, Göttingen 1968.
[53] Norton Simon, *A Uniqueness Proof for* F_1, in preparation.
[54] Norton Simon, unpublished work [*On "Norton algebras"*].

[55] Norton Simon, The Construction of J_4. In: *The Santa Cruz Conference on Finite Groups*, American Mathematical Society, Providence, 1980.

[56] O'Nan Michael, Automorphisms of Unitary Block Designs, *J. Algebra* **20** (1972), pp. 495–511.

[57] O'Nan Michael, A Characterization of $U_3(q)$, *J. Algebra* **22** (1972), pp. 254–296.

[58] Ree Rimhak, A Family of Simple Groups Associated with the Lie Algebra F_4, *Amer. J. Math.* **83** (1961), pp. 401–420.

[59] Ree Rimhak, A Family of Simple Groups Associated with the Lie Algebra G_2, *Amer. J. Math.* **83** (1961), pp. 432–462.

[60] Ronan Mark and Smith Steven D., 2-local Geometries for Finite Groups. In: *The Santa Cruz Conference on Finite Groups*, American Mathematical Society, Providence, 1980.

[61] Smith Stephen D., Nonassociative Commutative Algebras for Triple Covers of 3-transposition Groups, **24** (1977), pp. 273–289.

[62] Steinberg Robert, Variations on a Theme of Chevalley, *Pacific J. Math.* **9** (1959), pp. 875–891.

[63] Suzuki Michio, On a Class of Doubly Transitive Groups: I, II, *Ann. of Math.* **75** (1962), pp. 105–145; **79** (1964), pp. 514–589.

[64] Thompson John G., A Conjugacy Theorem for E_8, *J. Algebra* **38** (1976), pp. 525–530.

[65] Thompson John G., A Simple Subgroup of E_8 (3). In: N. Iwahori (ed.), *Finite Groups Symposium*, Japan Soc. for Promotion of Science, Tokyo, 1976, pp. 113–116.

[66] Thompson John G., The Uniqueness of the Fischer–Griess Monster, *Bull. London Math. Soc.* **11** (1979), pp. 340–346.

[67] Thompson John G., Some Numerology between the Fischer–Griess Monster and the Elliptic Modular Function, *Bull. London Math. Soc.* **11** (1979), pp. 352–353.

[68] Thompson John G., Towards a Characterization of $E_2^*(q)$: I, II, III, *J. Algebra* **7** (1967), pp. 406–414; **20** (1972), pp. 610–621; **49** (1977), pp. 163–166.

[69] Venkov B. B., On the Classification of Integral Even Unimodular 24 Dimensional Lattices, *Proc. Steklov Inst. Math.* 1980.

[70] Witt E., Die 5-fach transitiven Gruppen von Mathieu, *Abh. Math. Sem. Univ. Hamburg* **12** (1938), pp. 256–264.

Added in proof:

[71] Frenkel I., Lepowsky J., and Meurman A., to appear in *Proc. Nat. Acad. Sci. USA*.

[72] Griess Robert L., Jr., The Monster and Its Nonassociative Algebra, to appear in *Proceedings of the 1982 Montreal conference on finite groups*.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

MIKHAEL GROMOV

Infinite Groups as Geometric Objects

1.

Let a group Γ faithfully and isometrically act on a metric space X. We are interested in those algebraic properties of Γ which are reflected in the geometry of X and the action of Γ on X.

Examples. 1. A. The word-metric. If $\gamma_1, \ldots, \gamma_k$ is a finite system of generators of Γ, then there is a unique maximal left-invariant metric on Γ for which $\operatorname{dist}(e, \gamma_i^{\pm 1}) = 1$, $i = 1, \ldots, k$, where e is the identity element in Γ. This is called the *word-metric* and Γ acts on $X = (\Gamma, \operatorname{dist})$.

1.A′. If Γ is f.p. (finitely presented), then there exists a closed manifold V of any given dimension $n \geqslant 4$, such that $\Gamma \approx \pi_1(V)$. Indeed, take a 2-dimensional cell complex P for which $\pi_1(P, p_0) \approx \Gamma$, then embed $P \hookrightarrow \boldsymbol{R}^n$, $n \geqslant 5$, and take the boundary of a regular neighborhood of P in $\boldsymbol{R}^n$ for V. By fixing a Riemannian metric in V we obtain the induced metric on the universal cover $X' = \tilde{V}$ and Γ acts on X' by isometric deck transformations.

These examples are tied together with the following

1.B. Definitions. The *dilation* of a map $f: X \to X'$ between two metric spaces is

$$\operatorname{dil} f = \sup_{x,y} \operatorname{dist}\big(f(x), f(y)\big)/\operatorname{dist}(x, y).$$

The *distance* between two maps f_1 and f_2 is

$$\operatorname{dist}(f_1, f_2) = \sup_{x} \operatorname{dist}\big(f_1(x), f_2(x)\big).$$

A map $\tau: X \to X$ is called a *tranlation* if $\operatorname{dist}(\tau, \mathrm{id}) < \infty$. For instance, central elements in Γ are translations for the word-metric.

A map $f\colon X \to X'$ is called *v.L.* (*virtually Lipschitz*) if there exists a translation $\tau\colon X \to X$ such that $\operatorname{dil}(f \circ \tau) < \infty$. We call f *quasi-isometry* (*q.i.*) if there is a v.L. map $g\colon X' \to X$, for which $f o g$ and $g o f$ are translations.

Every isomorphism $f\colon \Gamma \to \Gamma'$ between finitely generated groups obviously is a quasi-isometry for arbitrary word-metrics in Γ and Γ'.

The orbit map $f = f_{\tilde{v}}\colon \Gamma \to \tilde{V}$ for $f\colon \gamma \to \gamma(\tilde{v})$ for the above deck action of $\Gamma = \pi_1(V)$ on $\tilde{V}$ clearly is a quasi-isometry for all $\tilde{v} \in V$ and for all compact manifolds V.

Let Γ and Γ' be discrete finitely generated subgroups in an arbitrary locally compact group G. Suppose Γ and Γ' are *cocompact* in G which means the existence of a compact subset $D \subset G$ such that $\Gamma \cdot D = \Gamma' \cdot D = G$. Then there exists a map $f\colon \Gamma \to \Gamma'$, such that $f(\gamma) \in \gamma \cdot D$ for all $\gamma \in \Gamma$, and every map f with this property clearly is a quasi-isometry $\Gamma \to \Gamma'$.

2.

Many (algebraic) properties of f.g. groups are q.i. invariants of the word-metric as the following examples show.

2.A. *If a torsion-free group Γ is quasi-isometric to a (nontrivial) free product, then Γ itself is such a product.*

Indeed, a famous theorem of Stallings [14] provides a *quasi-isometry invariant* description of free products:

A torsion free group Γ is a (nontrivial) free product if and only if Γ has infinitely many ends.

If Γ is finitely presented and so $\Gamma \approx \pi_1(V)$ for a compact manifold V, then ends of Γ correspond to ends of the universal covering $\tilde{V}$ of V and Stallings' theorem takes the following geometric form suggested by Matthew Brin [3].

2.A′. *If $\tilde{V}$ is disconnected at infinity, then there exists a closed hypersurface $H \subset \tilde{V}$ with the properties:*

(a) *The complement $\tilde{V} \setminus H$ contains at least two infinite components.*

(b) *If $\gamma H \subset \tilde{V}$ meets H for some $\gamma \in \Gamma$, then $\gamma H = H$.*

Proof of 2A′. The isometric cocompact action of Γ on $\tilde{V}$ insures the solvability of *Plateau's problem* in $\tilde{V}$. Thus one gets a hypersurface H of *minimal volume* which satisfies (a).

It is well known (and easy to prove) that no two codimension one (connected normally oriented and absolutely minimizing) solutions to Plateau's problem intersect unless they coincide (to grasp the idea look, for example, at minimal non-contractible closed geodesics in an infinite cyclic covering of the 2-torus T^2 with an arbitrary metric on T^2) and so H satisfies (b) as well.

Remark. This proof also applies if $\tilde{V}$ admits a (possibly noncompact) fundamental domain $\tilde{U} \subset \tilde{V}$ for which $\mathrm{Volume}_{n-1}(\partial \tilde{U}) < \infty$. Therefore the f.p. condition can be relaxed to f.g. Also notice that the above geometric proof can be translated to the algebraic language and then one comes back to the original argument of Stallings.

2.B. *If an f.g. group Γ is q.i. to a nilpotent (for example an abelian) group, then there is a nilpotent (correspondingly abelian) subgroup $\Gamma' \subset \Gamma$ for which Γ/Γ' is finite.*

Indeed, virtually nilpotent groups are characterized by the polynomial growth property (see [8]) which is q.i. invariant.

2.C. *If a torsion-free f.g. group Γ is q.i. to a cocompact lattice in $O(n, 1)$, $n \geqslant 2$, then Γ is isomorphic to such a lattice.*

The proof depends upon the conformal structure on the ideal boundary of Γ (see [12], [11], [9] and a forthcoming paper by Tukia) and it probably generalizes to arbitrary lattices in all semisimple Lie groups.

3. The Euler characteristic and cohomology with estimates

It is unknown whether the sign of the *virtual Euler characteristic* $\chi(\Gamma)$ (whenever defined, see [13]) is a q.i. invariant. However, $\chi(\Gamma)$ can be expressed by the L^2-cohomology of Γ (see [1], [5]) and vanishing of this cohomology is q.i. invariant. This leads to geometric criteria for vanishing (and sometimes nonvanishing) of $\chi(\Gamma)$ (see [4]).

Example. Let a *contractible* polyhedron X admit a cocompact isometric action of a group Γ and let $\tau_i\colon X \to X$, $i = 1, 2, \ldots$, be an infinite sequence of translations such that $\mathrm{dil}\,\tau_i \leqslant \mathrm{const} < \infty$ for all i and $\mathrm{dist}(\tau_i, \mathrm{id}) \to \infty$ for $i \to \infty$. Then one easily sees that the L^2-cohomology of X vanishes and so $\chi(\Gamma) = 0$ in case Γ is discrete. This generalizes a theorem of Gottlieb [7] (which was brought to my attention by S. Rossete) on the vanishing of $\chi(\Gamma)$ for groups with infinite centers.

In fact, it is usually more useful to have $\chi(\Gamma) \neq 0$, but no general geometric condition insures nonvanishing of the relevant L^2-cohomology. However, there is a closely related invariant, called the *simplicial volume*, which can be defined with *bounded cohomology* of Γ (see [10]) and which is known to be nonzero for certain *hyperbolic groups* (see below).

4. Hyperbolic spaces

A metric space X is called *hyperbolic* (*coarse hyperbolic* in the terminology of [9]) if there exist three positive constants C_1, C_2 and C_3 with the following properties:

(a) Take two arbitrary balls of radii R_1 and R_2 around some points x_1 and x_2 in X. Then the diameter of their intersection satisfies

$$\delta - C_1 \leqslant \operatorname{Diam}[B(x_1, R_1) \cap B(x_2, R_2)] \leqslant C_2(\delta + C_1).$$

where $\delta = R_1 + R_2 - \operatorname{dist}(x_1, x_2)$.

(b) Take three balls in X of radii R_i, $i = 1, 2, 3$, such that every two of them have nonempty intersection. Then the intersection of the three concentric balls of the respective radii $R_i + C_3$ is nonempty.

Here we are interested in hyperbolic spaces X with "large" isometry groups Γ operating on them. Since the hyperbolicity is a q.i. invariant (easy to show), the hyperbolicity of X amounts to that of Γ with the word-metric, in case of X/Γ compact.

Examples. 4.A. Let X be a *tree* with some "singular Riemannian" metric, such that the distance equals the length of the shortest path between two points. Such an X obviously is hyperbolic (for $C_1 = 0$, $C_2 = 1$ and $C_3 = 0$) and so every f.g. free group (with the word-metric) is hyperbolic.

4.B. *Noncompact symmetric spaces of rank* 1 *are hyperbolic.* In fact, the whole conception of hyperbolicity is an attempt to encompass the basic properties of symmetric spaces and their groups of isometries (in particular arithmetic and S-arithmetic groups) into a general geometric framework. Notice that the definition of hyperbolicity needs a nontrivial modification in order to include symmetric spaces of rank $\geqslant 2$ and their combinatorial counterparts which are called *Bruhat–Tits buildings*.

4.C. *Complete simply connected manifolds X of strictly negative curvature, $K \leqslant -\varepsilon < 0$, are hyperbolic.* One may even allow such an X to have a boundary, provided the boundary is convex. Furthermore, one may extend the notion of curvature to *spaces with singularities*, and then again $K \leqslant -\varepsilon < 0$ implies hyperbolicity (see [9]). Here are specific examples which explain the meaning of "singular negative curvature".

4.C′. Start with a complete noncompact manifold Y of constant curvature -1, such that $\operatorname{Vol} Y < \infty$. A submanifold C of full dimension in Y is called a *cusp* if the universal covering of C is (isometric to) a *horoball* B in the universal covering $\tilde{Y}$ of Y. (A *horoball* by definition is a union of an increasing family of balls of radii $R \to \infty$, $B = \bigcup_{R\to\infty} B(R)$, such that the boundary spheres $\partial B(R)$ have a common point for all R, i.e. $\bigcap_{R\to\infty} \partial B(R) \neq \emptyset$). It is well known that by chopping away finitely many disjoint cusps in Y one obtains a compact submanifold Y_0 of full dimension in Y whose boundary ∂Y_0 consists of finitely many smooth closed connected hypersurfaces $H_i \subset Y$, $i = 1, \ldots, k$ = the number of cusps in Y. (Notice, that every H_i is covered by a *horosphere* and so the induced metric in H_i is Riemannian flat for $i = 1, \ldots, k$.) Let us attach the unit cone to each H_i, $i = 1, \ldots, k$ (this can be done by taking first an isometric imbedding of each H_i into the unit sphere $S^{N-1} \subset \boldsymbol{R}^N$ for large N and then by taking ordinary cones from the center) and consider the resulting singular space X. One can show (with appropriate definitions) that this X has negative curvature under the following geometric condition

$$\text{every closed curve in } \partial Y_0 \text{ of length } \leqslant 2\pi \text{ is contractible.} \qquad (*)$$

Hence, the fundamental group $\pi_1(X)$ is infinite hyperbolic, provided $(*)$ is satisfied. (Since the fundamental group $\pi_1(Y_0) = \pi_1(Y)$ is *residually finite* condition $(*)$ is always satisfied for sufficiently large finite coverings of Y_0. Thus one obtains many manifolds of constant curvature -1 to which the above applies.)

4.C″. Fix integers $k \geqslant 6$ and $l \geqslant 2$. Then there exists a unique simply connected 2-dimensional polyhedron X whose all 2-cells are plane regular k-gons and such that

(a) The intersection of any two k-gons in X is (if nonempty) a common edge or a vertex.

(b) Every edge in X has l adjacent k-gons.

(c) Every vertex in X has $l+1$ adjacent edges (and hence $l(l+1)/2$ adjacent k-gons).

In other words, the *link* of every vertex in X is the complete graph with l vertices. This X has nonpositive curvature and $K < 0$ for $k \geqslant 7$. The isometry group is cocompact on X.

4.C″. There is a unique simply connected 3-dimensional polyhedron X whose 3-cells are regular dodecahedra and the link of each vertex is the 2-skeleton of the l-dimensional octahedron (for a given fixed $l = 3, 4, \ldots$). This X has $K < 0$ and the isometry group is cocompact. (If $l = 3$, one gets the hyperbolic 3-space $\boldsymbol{H}^3$ paved by dodecahedra).

4.D. Ramified covers. Take a totally geodesic submanifold V_0 of codimension 2 in a compact manifold V of negative curvature. There is a unique simply connected space X and a map $X \to V$ which is a covering over $V \setminus V_0$ and which *ramifies* at V_0 with a prescribed *ramification number* $\varrho = 1, 2, \ldots, \infty$. The induced singular metric in X has $K < 0$ and the isometry group is cocompact on X. If $\varrho = 1$, this is the ordinary universal covering $\tilde{V}$. If $\varrho = \infty$, then X is the (not locally compact) completion of the universal covering $\widetilde{V \setminus V_0}$. Furthermore, this construction applies to an arbitrary *locally convex subset* V_0 in V (compare [9]).

4.E. Small cancellation spaces. There are combinatorial criteria for hyperbolicity in terms of suitable covering of X by "small" subsets. These are systematically studied in the combinatorial group theory for the simplest case $\dim X = 2$. For example, $\frac{1}{6}$-groups are combinatorially hyperbolic and, hence, hyperbolic (for the word-metric). Many high-dimensional examples come from *Coxeter groups* (see [2], [6]). Let, for instance, V_0 be a compact oriented manifold of dimension n with a non-empty boundary. Then one can obtain (using an appropriate "sufficiently fine" triangulation of ∂V_0, see [6]) the following objects:

(i) A compact oriented n-dimensional manifold V *without boundary* and an embedding $a\colon V_0 \to V$ such that the fundamental group $\pi_1(V_0)$ *injects* into $\pi_1(V)$.

(ii) An aspherical compact oriented n-dimensional pseudomanifold W without boundary and a degree one map $b\colon V \to W$ such that the fundamental group $\pi_1(W)$ (being a finite index subgroup in a Coxeter group) is hyperbolic and the composed homomorphism $(b \circ a)_*\colon \pi_1(V_0) \to \pi_1(W)$ is zero.

(iii) If, furthermore, V_0 is aspherical, then V also is aspherical. Moreover, if $\pi_1(V_0)$ is hyperbolic, then $\pi_1(V)$ is hyperbolic as well.

4.F. There is a certain amount of surgery one can perform over hyperbolic spaces (like shrinking cusps in 4.C′, see [9]) and then starting with 4.A–4.E one has a huge amount of hyperbolic spaces and groups. The main problem is to classify them up to quasi-isometry. An important invariant of a hyperbolic space X is the ***ideal boundary*** ∂X, on which all isometries (and even quasi-isometries) of X act by homeomorphisms (see [9]). Most algebraic properties of hyperbolic groups (in particular, of small cancellation groups) Γ are immediate (see [9]) from (hyperbolic) dynamic behavior of the action of Γ on $\partial\Gamma$. There are, however, more subtle questions (such as evaluating the simplicial norm on $H_*(\Gamma)$, see [10], or the determination of all quasi-isometries of Γ modulo translations) which need additional ideas. Important results were announced (a private communication) by I. Rips (Jerusalem). His results would imply, for instance, nonexistence of a map $V \to V$ of degree $d \geqslant 2$, where V is an aspherical pseudomanifold without boundary with $\pi_1(V)$ hyperbolic (compare [10]). This, in fact, can be verified for the above examples with the techniques of [10].

5

In the limited space we could only mention a few geometric aspects in infinite groups. More examples can be found in the following references.

Finally, I wish to thank Nicolaas Kuiper for a careful reading of the first draft of this paper and for his many critical remarks and suggestions.

References

[1] Atiyah M. F., Elliptic Operators, Discrete Groups and von Neumann Algebras, Soc. Math. de France, *Astérisque* **32–33** (1976), pp. 43–72.

[2] Bourbaki N., XXXIV, *Groupes et algèbres de Lie*, Part 2, Hermann, Paris 1968.

[3] Brin M., Torsion Free Actions on 1-Acyclic Manifolds and the Loop Theorem, *Topology* **20** (1981), pp. 353–363.

[4] Cheeger J. and Gromov M., *Bounds on the von Neumann Dimension of L^2-Cohomology and the Gauss–Bonnet Theorem for Open Manifolds*, to appear.

[5] Cohen J., Von Neumann Dimension and the Homology of Covering Spaces, *Quart. J. Math. Oxford* **30** (1979), pp. 133–142.

[6] Davis M., Groups Generated by Reflections and Aspherical Manifolds Not Covered by Euclidean Space, *Ann. of Math.* **117** (2) (1983), pp. 293–325.

[7] Gottlieb D., A Certain Subgroup of the Fundamental Group, *Amer. J. Math.* **87** (1965), pp. 840–865.

[8] Gromov M., Groups of Polynomial Growth and Expanding Maps, *Publ. Math. IHES* **53** (1981), pp. 53–78.

[9] Gromov M., Hyperblic Manifolds, Groups and Actions. In: *Riemannian Surfaces and Related Topics*, Ann. Math. Studies 97, (1981), pp. 183–215.

[10] Gromov M., Volume and Bounded Cohomology, *Publ. Math. IHES* **56** (1983), pp. 213–307.

[11] Margulis G., The Isometry of Closed Manifolds of Constant Negative Curvature with the Same Fundamental Group, *Dokl. Akad. Nauk SSSR* **192** (1970), pp. 736–737.
[12] Mostow G., *Strong Rigidity of Locally Symmetric Spaces*, Ann. Math. Studies 78 (1973), Princeton.
[13] Serre J. P., *Cohomologie des groupes discrets*, Prosp. in Math., Ann. Math. Studies 70 (1971), Princeton.
[14] Stallings J., On Torsion Free Groups with Infinitely Many Ends, *Ann. of Math.* **88** (1968), pp. 312–334.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

J. C. JANTZEN

Einhüllende Algebren halbeinfacher Lie-Algebren*

1. Es sei $\mathfrak{g}$ eine endlich dimensionale, komplexe Lie-Algebra. Wir bezeichnen ihre universelle einhüllende Algebra mit $U(\mathfrak{g})$ und deren Zentrum mit $Z(\mathfrak{g})$. Auf dem letzten Kongreß hat J. Dixmier [9] erläutert, warum man sich für die Struktur gerade dieser Ringe interessiert, insbesondere für den Raum $\mathscr{X}$ der primitiven Ideale in $U(\mathfrak{g})$, und einen Überblick über die bis dahin gefundenen Ergebnisse gegeben. Hier möchte ich über neuere Entwicklungen berichten, wenngleich gewisse Überschneidungen mit [9] unvermeidlich sind. Ich werde von nun an annehmen, daß $\mathfrak{g}$ halbeinfach ist, und will vor allem Zusammenhänge mit der Darstellungstheorie von $\mathfrak{g}$ und von $\mathfrak{g}\times\mathfrak{g}$ ansprechen, während A. Joseph in seinem Vortrag auf diesem Kongreß voraussichtlich mehr die ringtheoretischen Aspekte der Theorie betonen wird und auch auf den Fall, daß $\mathfrak{g}$ beliebig ist, eingehen wird.

2.1. Ein Ideal in $U(\mathfrak{g})$ ist offensichtlich dasselbe wie ein Unterbimodul für die natürliche Struktur von $U(\mathfrak{g})$ als $\big(U(\mathfrak{g}), U(\mathfrak{g})\big)$-Bimodul. Wir fassen stets einen $\big(U(\mathfrak{g}), U(\mathfrak{g})\big)$-Bimodul X auch als $(\mathfrak{g}\times\mathfrak{g})$-Modul auf (und umgekehrt): Für $(a, b)\in\mathfrak{g}\times\mathfrak{g}$ und $x\in X$ setzt man $(a, b)x = ax - xb$.

In $\mathfrak{g}\times\mathfrak{g}$ ist $\mathfrak{k} = \{(a, a) \mid a\in\mathfrak{g}\}$ eine Unteralgebra, die wir stets unter $a\mapsto(a, a)$ mit $\mathfrak{g}$ identifizieren. Bei der Operation von $\mathfrak{g}\times\mathfrak{g}$ auf $U(\mathfrak{g})$ oben gilt $\dim U(\mathfrak{k})x < \infty$ für alle $x\in U(\mathfrak{g})$, d.h. daß $U(\mathfrak{g})$ ein lokal endlicher $U(\mathfrak{k})$-Modul ist. Wir wollen hier allgemein einen $(\mathfrak{g}\times\mathfrak{g})$-Modul X einen Harish-Chandra-Modul nennen, wenn er (a) lokal endlich als $U(\mathfrak{k})$-Modul ist und wenn (b) $\dim_C \mathrm{Hom}_{\mathfrak{k}}(E, X) < \infty$ für alle endlich dimensionalen $\mathfrak{k}$-Moduln E gilt. Für ein Ideal I von $U(\mathfrak{g})$ erfüllt $U(\mathfrak{g})/I$ stets (a), aber (b) nur dann, wenn $I\cap Z(\mathfrak{g})$ endliche Kodimension in $Z(\mathfrak{g})$ hat (vgl. [8], chap. 8).

---

* Auf dem Kongreß wurde der Vortrag unter dem Titel "Enveloping algebras of semisimple Lie algebras" auf English gehalten.

2.2. Für jedes $I \in \mathscr{X}$ gibt es einen zentralen Charakter $\chi: Z(\mathfrak{g}) \to C$ von $U(\mathfrak{g})$ mit $I \cap Z(\mathfrak{g}) = \mathrm{Kern}(\chi)$; nach 2.1 ist $U(\mathfrak{g})/I$ daher ein Harish-Chandra-Modul. Dies gilt auch für $U(\mathfrak{g})/U(\mathfrak{g})\mathrm{Kern}(\chi)$; auf diesem Modul operiert nun $Z(\mathfrak{g} \times \mathfrak{g}) \simeq Z(\mathfrak{g}) \otimes Z(\mathfrak{g})$ durch einen Charakter. Aus Sätzen von Harish-Chandra folgt, daß jeder Harish-Chandra-Modul mit einem zentralen Charakter endliche Länge hat. (Ein rein algebraischer Beweis steht in [4].) Man findet also alle primitiven Ideale unter den Untermoduln der $(\mathfrak{g} \times \mathfrak{g})$-Moduln endlicher Länge $U(\mathfrak{g})/U(\mathfrak{g})\mathrm{Kern}\,\chi$, wobei χ die Homomorphismen $Z(\mathfrak{g}) \to C$ durchläuft.

2.3. Für jeden $(U(\mathfrak{g}), U(\mathfrak{g}))$-Bimodul X bezeichnen wir mit $\mathrm{LAnn}(X) = \{u \in U(\mathfrak{g}) \mid ux = 0$ für alle $x \in X\}$ seinen Linksannullator und analog mit $\mathrm{RAnn}(X)$ seinen Rechtsannullator. Für ein Ideal I von $U(\mathfrak{g})$ gilt offensichtlich $I = \mathrm{LAnn}(U(\mathfrak{g})/I) = \mathrm{RAnn}(U(\mathfrak{g})/I)$. Für primitives I hat $U(\mathfrak{g})/I$ endliche Länge; da I prim ist, muß es daher Kompositionsfaktoren X und X' von $U(\mathfrak{g})/I$ mit

$$I = \mathrm{LAnn}(X) = \mathrm{RAnn}(X') \tag{1}$$

geben. (Man kann sogar für X und X' den Sockel von $U(\mathfrak{g})/I$ nehmen [10].) Eine Klassifikation der einfachen Harish-Chandra-Moduln führt also zu einer Beschreibung von $\mathscr{X}$.

2.4. Es sei X ein Harish-Chandra-Modul, der endlich erzeugbar über $U(\mathfrak{g} \times \mathfrak{g})$ ist. Wegen 2.1 (a) kann man einen endlich dimensionalen Teilraum $E \subset X$ mit $X = U(\mathfrak{g} \times \mathfrak{g})E$ und $U(\mathfrak{k})E = E$ finden. Dann folgt $X = U(\mathfrak{g})E = EU(\mathfrak{g})$; somit ist X als Links- und Rechtsmodul über $U(\mathfrak{g})$ endlich erzeugbar, also auch noethersch. Wir erhalten einen surjektiven Homomorphismus [27]

$$U(\mathfrak{g})/\mathrm{RAnn}(X) \twoheadrightarrow E^l \otimes X, \tag{2}$$

von $(\mathfrak{g} \times \mathfrak{g})$-Moduln, wenn wir ein $(a, b) \in \mathfrak{g} \times \mathfrak{g}$ auf E so wie $(a, a) \in \mathfrak{k}$ operieren lassen. (Daß wir diese Operation betrachten, soll durch die Notation E^l ausgedrückt werden.) Tensorieren wir die Abbildung in (2) mit E^* und betten C als $C\mathrm{id}_E$ in $\mathrm{End}_C(E) = E^* \otimes E$ ein, so erhalten wir einen injektiven Homomorphismus von $(\mathfrak{g} \times \mathfrak{g})$-Moduln [27]

$$U(\mathfrak{g})/\mathrm{RAnn}(X) \hookrightarrow X \otimes (E^*)^l. \tag{3}$$

Für zwei endlich erzeugbare Harish-Chandra-Moduln X, Y folgt nun [27], daß $\mathrm{RAnn}(X) \subset \mathrm{RAnn}(Y)$ genau dann gilt, wenn es einen endlich dimensionalen $\mathfrak{k}$-Modul E gibt, sodaß Y zu einem Subquotienten von $X \otimes E^l$ isomorph ist. (Gilt $\mathrm{RAnn}(X) \subset \mathrm{RAnn}(Y)$, so hat man eine Surjektion $U(\mathfrak{g})/\mathrm{RAnn}(X) \twoheadrightarrow U(\mathfrak{g})/\mathrm{RAnn}(Y)$ und wendet (2) auf Y sowie (3) auf X an; die Umkehrung ist trivial.)

3.1. Nach 2.3/4 führt eine Klassifikation der einfachen Harish-Chandra-Moduln und die Bestimmung der Kompositionsfaktoren in deren Tensorprodukten mit endlich dimensionalen Darstellungen zu einer Beschreibung von $\mathscr{X}$ als geordneter Menge. Die einfachen Harish-Chandra-Moduln hat zuerst D. P. Želobenko klassifiziert. Mit anderen Methoden geschieht dies in [4] und [20], denen ich hier folge, sowie in [11].

3.2. Für $\mathfrak{g}$-Moduln M, N ist $\mathrm{Hom}_C(M, N)$ unter $\big((a, b)\varphi\big)(m) = a\varphi(m) - \varphi(bm)$ für alle $a, b \in \mathfrak{g}$, $\varphi \in \mathrm{Hom}_C(M, N)$, $m \in M$ ein $(\mathfrak{g} \times \mathfrak{g})$-Modul. Dann bilden die $\varphi \in \mathrm{Hom}_C(M, N)$ mit $\dim_C U(\mathfrak{k})\varphi < \infty$ einen $(\mathfrak{g} \times \mathfrak{g})$-Untermodul $\mathscr{L}(M, N)$ von $\mathrm{Hom}_C(M, N)$. Nach Konstruktion erfüllt $\mathscr{L}(M, N)$ die Bedingung 2.1 (a); um 2.1 (b) nachzuprüfen, muß man

$$\mathrm{Hom}_{\mathfrak{k}}\big(E, \mathscr{L}(M, N)\big) = \mathrm{Hom}_{\mathfrak{k}}\big(E, \mathrm{Hom}_C(M, N)\big) \simeq \mathrm{Hom}_{\mathfrak{g}}(E \otimes M, N)$$
$$\simeq \mathrm{Hom}_{\mathfrak{g}}(M, E^* \otimes N) \tag{4}$$

für alle endlich dimensionalen $\mathfrak{g}$-Moduln E untersuchen. Man sieht leicht, daß $\mathscr{L}$ ein Funktor ist, linksexakt in beiden Argumenten, und daß $\mathscr{L}(M, M)$ für jeden $\mathfrak{g}$-Modul M ein Ring ist, wenn man die Verknüpfung von Abbildungen als Multiplikation nimmt.

3.3. Haben die $\mathfrak{g}$-Moduln M, N endliche Länge, so ist $\mathscr{L}(M, N)$ ein Harish-Chandra-Modul endlicher Länge ([22], III). (Wegen der Linksexaktheit von $\mathscr{L}$ muß man nur einfache M und N betrachten. Da dann $Z(\mathfrak{g} \times \mathfrak{g})$ durch einen Charakter auf $\mathscr{L}(M, N)$ operiert, muß man nach 2.2 nur zeigen, daß $\mathscr{L}(M, N)$ die Bedingung (b) in 2.1 erfüllt. Dies folgt aus (4) und elementaren Eigenschaften der Gel'fand–Kirillov-Dimension sowie der Multiplizität (auch Bernstein-Grad genannt) von $\mathfrak{g}$-Moduln.

3.4. Aus diesem Satz und den Bemerkungen zu Anfang von 2.4 folgt für jeden $\mathfrak{g}$-Modul M endlicher Länge, daß der Ring $\mathscr{L}(M, M)$ noethersch ist. In diesen Ring ist $U(\mathfrak{g})/\mathrm{Ann}(M)$ als Teilring eingebettet. Kostant hatte gefragt, ob diese Einbettung für einfaches M ein Isomorphismus ist. Zwar gilt dies im allgemeinen nicht [7], [23], doch hat man in vielen Fällen immerhin einen Isomorphismus der totalen Quotientenringe [19].

3.5. Ein Vergleich mit $U(\mathfrak{g})/\mathrm{Ann}(M)$ ist in jedem Fall wünschenswert, weil sich die Struktur von $\mathscr{L}(M, M)$ häufiger bestimmen läßt als die von $U(\mathfrak{g})/\mathrm{Ann}(M)$. Man nehme zum Beispiel eine parabolische Unteralgebra $\mathfrak{p}$ von $\mathfrak{g}$ sowie einen einfachen, endlich dimensionalen $\mathfrak{p}$-Modul E und bilde den induzierten $\mathfrak{g}$-Modul $M_{\mathfrak{p}}(E) = U(\mathfrak{g}) \underset{U(\mathfrak{p})}{\otimes} E$. Dann ist $\mathscr{L}\big(M_{\mathfrak{p}}(E), M_{\mathfrak{p}}(E)\big)$ prim; sein totaler Quotientenring ist ein Matrixring vom Grad $\dim(E)$ über einem Weylschen Schiefkörper [16], [18]. Dagegen braucht $U(\mathfrak{g})/\mathrm{Ann}\big(M_{\mathfrak{p}}(E)\big)$ nicht prim zu sein; ist es dieses doch, so hat man keine so präzise Information über den Quotientenring wie eben.

4.1. Um alle einfachen Harish-Chandra-Moduln zu klassifizieren, führen wir einige Notationen ein. Es seien $\mathfrak{b}$ eine Borel-Unteralgebra und $\mathfrak{h} \subset \mathfrak{b}$ eine Cartan-Unteralgebra von $\mathfrak{g}$. Wir bezeichnen das Wurzelsystem von $\mathfrak{g}$ mit $R = R(\mathfrak{g}, \mathfrak{h})$ und die Menge der positiven Wurzeln mit $R^+ = R(\mathfrak{b}, \mathfrak{h})$, die durch R^+ festgelegte Basis von R mit B und die halbe Summe der positiven Wurzeln mit ϱ. Es sei W die Weylgruppe von $\mathfrak{g}$ relativ $\mathfrak{h}$; für alle $\lambda \in \mathfrak{h}^*$ und $w \in W$ schreiben wir $w \cdot \lambda = w(\lambda + \varrho) - \varrho$.

4.2. Jedes $\lambda \in \mathfrak{h}^*$ läßt sich durch $\lambda([\mathfrak{b}, \mathfrak{b}]) = 0$ zu einer eindimensionalen Darstellung von $\mathfrak{b} = \mathfrak{h} \oplus [\mathfrak{b}, \mathfrak{b}]$ fortsetzen. Es sei $M(\lambda)$ der davon induzierte $\mathfrak{g}$-Modul; dies ist der Verma-Modul mit höchstem Gewicht λ. Der Faktormodul $L(\lambda) = M(\lambda)/\mathrm{rad}\, M(\lambda)$ ist einfach. (Man beachte, daß diese Moduln häufig als $M(\lambda + \varrho)$ und $L(\lambda + \varrho)$ bezeichnet werden.) Auf jedem $M(\lambda)$ operiert $Z(\mathfrak{g})$ durch einen Charakter, den wir mit χ_λ bezeichnen wollen; für $\lambda, \mu \in \mathfrak{h}^*$ ist $\chi_\lambda = \chi_\mu$ zu $\mu \in W \cdot \lambda$ äquivalent. Jedes $M(\lambda)$ hat endliche Länge; seine Kompositionsfaktoren haben die Form $L(\mu)$ mit $\mu \in W \cdot \lambda$ (vgl. [8], chap. 7). Die Multiplizität von $L(\mu)$ als Kompositionsfaktor in einem $\mathfrak{g}$-Modul M endlicher Länge bezeichnen wir mit $[M: L(\mu)]$.

4.3. Für alle $\lambda \in \mathfrak{h}^*$ ist $I(\lambda) = \mathrm{Ann}\big(L(\lambda)\big)$ ein primitives Ideal in $U(\mathfrak{g})$. Nach [4], [20] ist jeder einfache Harish-Chandra-Modul zu einem $\mathscr{L}\big(M(\mu), L(\lambda)\big)$ mit $\lambda, \mu \in \mathfrak{h}^*$ isomorph. Für $\mathscr{L}\big(M(\mu), L(\lambda)\big) \neq 0$ gilt offensichtlich $\mathrm{LAnn}\, \mathscr{L}\big(M(\mu), L(\lambda)\big) = I(\lambda)$, also erhalten wir den Satz von Duflo [10]:

$$\mathscr{X} = \{I(\lambda) \mid \lambda \in \mathfrak{h}^*\}. \tag{5}$$

Die Fasern der Abbildung $I \mapsto I \cap Z(\mathfrak{g})$ von $\mathscr{X}$ auf das maximale Spektrum von $Z(\mathfrak{g})$ sind daher die $\mathscr{X}_\lambda = \{I(w \cdot \lambda) \mid w \in W\}$ mit $\lambda \in \mathfrak{h}^*$.

4.4. Für alle $\alpha \in R$ sei $\alpha^\vee$ die duale Wurzel; wir setzen $P(R) = \{\lambda \in \mathfrak{h}^* \mid \langle \lambda, \alpha^\vee \rangle \in \mathbf{Z}$ für alle $\alpha \in R\}$. Die Struktur von $\mathscr{X}_\lambda$ hängt nur davon ab, für welche $\alpha \in R$ die Zahl $\langle \lambda, \alpha^\vee \rangle$ zu $\boldsymbol{Z}$ gehört und wann sie gleich 0 ist [14]. Wir beschränken uns hier auf den Fall $\lambda \in P(R)$, wo die $\mathscr{X}_\lambda$ am kompliziertesten, die Notationen am einfachsten sind. Dazu betrachten wir die Kategorie $\mathcal{O}$ der $\mathfrak{g}$-Moduln endlicher Länge, deren Kompositionsfaktoren alle die Form $L(\mu)$ mit $\mu \in P(R)$ haben und die als $\mathfrak{h}$-Moduln gleich der direkten Summe ihrer Gewichtsräume sind. (Dies ist nur ein Teil der in [3] eingeführten Kategorie.)

4.5. Wir bezeichnen die Menge der dominanten Gewichts mit $P(R)^{++} = \{\mu \in \mathfrak{h}^* \mid \langle \mu, \alpha^\vee \rangle \in \boldsymbol{N}$ für alle $\alpha \in R^+\}$. Dann ist $-\varrho + P(R)^{++}$ ein Fundamentalbereich für die (um ϱ verschobene) Operation von W auf $P(R)$.

Für $\lambda \in P(R)^{++}$ ist $M(\lambda)$ ein projektives Objekt [3] in $\mathcal{O}$; daher ist der Funktor $M \rightsquigarrow \mathscr{L}(M(\lambda), M)$ von $\mathcal{O}$ in die Kategorie der Harish-Chandra-Moduln exakt. Man kann sogar zeigen, daß er eine Äquivalenz von Kategorien zwischen $\mathcal{O}$ und der Kategorie der Harish-Chandra-Moduln X endlicher Länge mit $\mathrm{Kern}(\chi_\lambda) \subset \mathrm{RAnn}(X)$ induziert [4]. Für alle $\mu \in P(R)$ ist $\mathscr{L}(M(\lambda), L(\mu))$ einfach [4], [20].

Es seien $\lambda, \mu \in P(R)^{++}$ und $w \in W$. Während nun die Gleichung $\mathrm{LAnn}\,\mathscr{L}(M(\lambda), L(w \cdot \mu)) = I(w \cdot \mu)$ klar ist, benötigt man für $\mathrm{RAnn}\,\mathscr{L}(M(\lambda), L(w \cdot \mu)) = I(w^{-1} \cdot \lambda)$ mehr Theorie [17]. Man zeigt leicht, daß $\mathscr{L}(M(\lambda), M \otimes E) \simeq \mathscr{L}(M(\lambda), M) \otimes E^t$ für alle M in $\mathcal{O}$ und alle endlich dimensionalen $\mathfrak{g}$-Moduln E gilt. Aus 2.4 folgt nun leicht: Für $w, w' \in W$ ist $I(w^{-1} \cdot \lambda) \subset I(w'^{-1} \cdot \lambda)$ dazu äquivalent, daß es einen endlich dimensionalen $\mathfrak{g}$-Modul E mit $[L(w \cdot \mu) \otimes E : L(w' \cdot \mu)] \neq 0$ gibt [27], [1].

5.1. Nach den letzten Sätzen ist klar, daß man $\mathscr{X}_\lambda$ für $\lambda \in P(R)$ mit einer vollständigen Kenntnis der Kategorie $\mathcal{O}$ beschreiben kann. Wir betrachten daher $\mathcal{O}$ genauer und bilden insbesondere ihre Grothendieck-Gruppe. Für jedes M in $\mathcal{O}$ sei $[M]$ seine Klasse in dieser Gruppe; es gilt also $[M] = \sum\limits_{\mu} [M : L(\mu)][L(\mu)]$. Neben den $[L(\mu)]$ mit $\mu \in P(R)$ bilden auch die $[M(\mu)]$ mit $\mu \in P(R)$ eine Basis dieser Gruppe. Für alle $\lambda \in P(R)^{++}$ und $w, w' \in W$ gibt es $a_{w,w'}(\lambda) \in \boldsymbol{Z}$ mit $[L(w \cdot \lambda)] = \sum\limits_{w' \in W} a_{w,w'}(\lambda)[M(w' \cdot \lambda)]$.

5.2. Für jeden $\mathfrak{g}$-Modul M und alle $\mu \in \mathfrak{h}^*$ ist $M_\mu = \{m \in M \mid$ für alle $z \in Z(\mathfrak{g})$ gibt es $n \in \boldsymbol{N}$ mit $(z - \chi_\mu(z)1)^n m = 0\}$ ein Untermodul von M; gehört M zu $\mathcal{O}$, so ist M die direkte Summe der M_μ mit $\mu \in -\varrho + P(R)^{++}$. Für alle $\lambda, \mu \in -\varrho + P(R)^{++}$ definieren wir einen exakten Funktor T_λ^μ von $\mathcal{O}$ in sich durch $T_\lambda^\mu(M) = (M_\lambda \otimes E)_\mu$, wobei wir für E den endlich dimensionalen, einfachen $\mathfrak{g}$-Modul mit höchstem Gewicht in $W(\mu - \lambda)$ nehmen.

Für $w \in W$ setzen wir $\tau(w) = \{\alpha \in B \mid w(\alpha) \in -R^+\}$, für $\mu \in -\varrho + P(R)^{++}$ setzen wir $B_\mu^0 = \{\alpha \in B \mid \langle \mu + \varrho, \alpha^\vee \rangle = 0\}$. Für alle $\lambda \in P(R)^{++}$ und $\mu \in -\varrho + +P(R)^{++}$ gilt nun $T_\lambda^\mu M(w \cdot \lambda) \simeq M(w \cdot \mu)$ für alle $w \in W$, und $T_\lambda^\mu L(w \cdot \lambda) \simeq L(w \cdot \lambda)$ für $\tau(w) \supset B_\mu^0$, aber $T_\lambda^\mu L(w \cdot \lambda) = 0$ sonst [13]. Insbesondere folgt $a_{w,w'}(\lambda) = a_{w,w'}(\mu)$ für alle $\lambda, \mu \in P(R)^{++}$ und $w, w' \in W$; wir schreiben kurz $a_{w,w'}$ für diese Zahl.

5.3. Die $a_{w,w'}$ werden in [2] und [6] bestimmt. Es gilt $a_{w,w'} = \det(ww') \times \times P_{w_0 w', w_0 w}(1)$, wobei $w_0 \in W$ das Element mit $w_0(R^+) = -R^+$ ist. Ferner ist $P_{w,w'}$ für alle $w, w' \in W$ ein in [24] konstruiertes Polynom in einer Veränderlichen T, das sich durch Iteration berechnen läßt. Es gilt auch $[M(w \cdot \lambda) \colon L(w' \cdot \lambda)] = P_{w,w'}(1)$ für alle $\lambda \in P(R)^{++}$ und $w, w' \in W$.

Für alle $w \in W$ sei $l(w)$ die Anzahl der $\alpha \in R^+$ mit $w(\alpha) \notin R^+$. In [14] wird in jedem Verma-Modul $M(\mu)$ eine endliche Filtrierung $M(\mu) = M(\mu)_0 \supset M(\mu)_1 \supset \ldots$ konstruiert; wir setzen $M_i(\mu) = M(\mu)_i / M(\mu)_{i+1}$. Nach unveröffentlichten Arbeiten von Beilinson und Bernstein gilt nun [25]

$$P_{w,w'} = \sum_{j \geq 0} [M_j(w \cdot \lambda) \colon L(w' \cdot \lambda)] T^{(l(w') - l(w) - j)/2}. \tag{6}$$

Man benutzt dazu die Ergebnisse von [12] und zeigt auch [28]

$$P_{w,w'} = \sum_{j \geq 0} \dim_C \operatorname{Ext}^j_{\mathcal{O}}\big(M(w \cdot \lambda), L(w' \cdot \lambda)\big) T^{(l(w') - l(w) - j)/2}. \tag{7}$$

5.4. In 5.2 hängt $\operatorname{Ann}(T_\lambda^\mu M)$ nur von $\operatorname{Ann}(M)$ ab; daher erhält man so auch eine Abbildung auf gewissen Idealen von $U(\mathfrak{g})$. Für $\lambda \in P(R)^{++}$ kann man eine Abbildung τ von $\mathscr{X}_\lambda$ in die Potenzmenge von B durch $\tau(I) = \tau(w)$ für alle w mit $I = I(w \cdot \lambda)$ definieren. Ist dann $\mu \in -\varrho + P(R)^{++}$, so ist $\mathscr{X}_\mu$ als geordnete Menge zu $\{I \in \mathscr{X}_\lambda \mid \tau(I) \supset B_\mu^0\}$ unter $I(w \cdot \mu) \mapsto I(w \cdot \lambda)$ für $\tau(w) \supset B_\mu^0$ isomorph [5].

Es sei $\lambda \in P(R)^{++}$. Man kann einem $I \in \mathscr{X}_\lambda$ neben $\tau(I)$ auch eine verallgemeinerte τ-Invariante zuordnen [26], mit deren Hilfe man elementar zeigen kann, daß die Abbildung $w \mapsto I(w \cdot \lambda)$ für $\mathfrak{g} = \mathfrak{sl}_n$ eine Bijektion $\{w \in W \mid w^2 = 1\} \to \mathscr{X}_\lambda$ ist [26], [15], ein Ergebnis das zuerst mit komplizierten Hilfsmitteln in [18], II bewiesen wurde.

6.1. Nach 4.5 ist klar, daß wir die $[L(w \cdot \lambda) \otimes E \colon L(w' \cdot \lambda)]$ mit $\mu \in P(R)^{++}$ und $w, w' \in W$ kennen müssen. Lassen wir E die endlich dimensionalen $\mathfrak{g}$-Moduln durchlaufen, so erzeugen die $[(L(w \cdot \mu) \otimes E)_\mu]$ in der Grothendieck-Gruppe von $\mathcal{O}$ dieselbe Untergruppe wie die $\sum_{w' \in W} a_{w,w'} [M(w' w_1 \cdot \mu)]$ mit $w_1 \in W$.

Wir setzen $a_w = \sum_{w' \in W} a_{w,w'} w'^{-1} \in C[W]$. Die a_w bilden eine Basis von $C[W]$; für alle $f \in C[W]$ gibt es daher eindeutig bestimmte $[f: a_w] \in C$ mit $f = \sum_{w \in W} [f: a_w] a_w$. Es ist dann $\sum_{w' \in W} a_{w,w'} [M(w' w_1 \cdot \mu)] = \sum_{w' \in W} [w_1^{-1} a_w: a_{w'}][L(w' \cdot \mu)]$. Daher gilt $I(w^{-1} \cdot \mu) \subset I(w'^{-1} \cdot \mu)$ genau dann, wenn es ein $w_1 \in W$ mit $[w_1 a_w: a_{w'}] \neq 0$ gibt [21], [27]. Da stets $a_{w,w'} = a_{w^{-1}, w'^{-1}}$ ist [20], [4], gilt $I(w \cdot \mu) \subset I(w' \cdot \mu)$ genau dann, wenn es ein $w_1 \in W$ mit $[a_w w_1: a_{w'}] \neq 0$ gibt.

6.2. Es sei $\mu \in P(R)^{++}$. Für jedes $I \in \mathscr{X}_\mu$ bilden nach 6.1 die a_w mit $I \subset I(w \cdot \mu)$ die Basis eines $C[W]$-Rechtsuntermoduls V_I von $C[W]$. Teilt man V_I durch die Summe der V_J mit $J \in \mathscr{X}_\mu$ und $J \not\supseteq I$, so erhält man einen Faktormodul $\overline{V}_I$, der als Basis die Restklassen der a_w mit $I = I(w \cdot \mu)$ hat. Dann ist $C[W]$ zur direkten Summe der $\overline{V}_I$ mit $I \in \mathscr{X}_\mu$ isomorph. Für $\mathfrak{g} = \mathfrak{sl}_n$ ist jedes $\overline{V}_I$ einfach, im allgemeinen ist die Situation komplizierter [23], [1].

Man nennt für ein $w \in W$ die Menge $\{w' \in W \mid I(w \cdot \mu) = I(w' \cdot \mu)\}$ die linke Zelle von w. Diese Zellen bilden eine disjunkte Zerlegung von W; dem entspricht eine direkte Zerlegung der regulären Darstellung von W (in die $\overline{V}_{I(w \cdot \mu)}$). Dabei gehört zu jeder Zelle eine Darstellung, deren Grad die Mächtigkeit der Zelle ist. Die Einteilung in die Zellen kann man mit Hilfe der $P_{w,w'}$ leicht beschreiben [24].

6.3. Aus 6.1 und den bekannten Eigenschaften der $P_{w,w'}$ folgt, daß es einen involutorischen Antiautomorphismus σ von $\mathscr{X}_\mu$ mit $\sigma(I(w \cdot \mu)) = I(w w_0 \cdot \mu)$ für alle $w \in W$ gibt [22] III, [1]. Es ist dann $\overline{V}_{\sigma(I)}$ kanonisch zu $\overline{V}_I^* \otimes \det$ für alle $I \in \mathscr{X}_\mu$ isomorph [1].

6.4. Der Zusammenhang zwischen primitiven Idealen und der Darstellungstheorie der Weylgruppe wird vertieft, wenn man Goldie-Rang-Polynome und assoziierte Varietäten betrachten. Dazu möchte ich auf den Vortrag von A. Joseph auf diesem Kongreß verweisen.

Literatur

(Ausführlichere Angaben findet man in [15])

[1] Barbasch D., Vogan D., Primitive Ideals and Orbital Integrals in Complex Exceptional Groups, *J. Algebra* **80** (1983), S. 350–382.

[2] Beilinson A., Bernstein J., Localisation de $\mathfrak{g}$-modules, *C. R. Acad. Sci. Paris* (I) **292** (1981), S. 15–18.

[3] Bernštein I. N., Gel'fand I. M., Gel'fand S. I., Category of $\mathfrak{g}$-Modules, *Funct. Anal. Appl.* **10** (1976), S. 87–82.
[4] Bernstein J. N., Gelfand S. I., Tensor Products of Finite and Infinite Dimensional Representations of Semisimple Lie Algebras, *Compositio math.* **41** (1980), S. 245–285.
[5] Borho W., Jantzen J. C., Über primitive Ideale in der Einhüllenden einer halbeinfachen Lie-Algebra, *Invent. math.* **39** (1977), S. 1–53.
[6] Brylinski J. L., Kashiwara M., Kazhdan–Lusztig Conjecture and Holonomic Systems, *Invent. math.* **64** (1981), S. 387–410.
[7] Conze-Berline N., Duflo M., Sur les représentations induites des groupes semisimples complexes, *Compositio math.* **34** (1977), S. 307–336.
[8] Dixmier J., *Algèbres enveloppantes*, Paris–Bruxelles–Montréal, 1974.
[9] Dixmier J., Algèbres enveloppantes. In: *Proceedings of the International Congress of Mathematicians*, Helsinki, 1978, S. 693–698.
[10] Duflo M., Sur la classification des idéaux primitifs dans l'algébre enveloppante d'une algèbre de Lie semi-simple, *Ann. of Math.* **105** (1977), S. 107–120.
[11] Enright T. J., Lectures on Representations of Complex Semi-Simple Lie Groups, *Tata Institute Lectures*, Berlin–Heidelberg–New York, 1981.
[12] Gabber O., Joseph A., Towards the Kazhdan–Lusztig Conjecture, *Ann. scient. Éc. Norm. Sup.* **14** (4) (1981), S. 261–302.
[13] Jantzen J. C., Zur Charakterformel gewisser Darstellungen halbeinfacher Gruppen und Lie-Algebren, *Math. Z.* **140** (1974), S. 127–149.
[14] Jantzen J. C., Moduln mit einem höchsten Gewicht, *Lecture Notes in Mathematics* **750**, Berlin–Heidelberg–New York, 1979.
[15] Jantzen J. C., Einhüllende Algebren halbeinfacher Lie-Algebren, *Ergebnisse der Mathematik*, 3. Folge, Band 3, Berlin–Heidelberg–New York, 1983.
[16] Joseph A., On the Gel'fand–Kirillov Conjecture for Induced Ideals in the Semisimple Case, *Bull. Soc. math. France* **107** (1979), S. 139–159.
[17] Joseph A., On the Annihilators of the Simple Subquotients of the Principal Series, *Ann. scient. Éc. Norm. Sup.* (4) **10** (1977), S. 419–440.
[18] Joseph A., Towards the Jantzen Conjecture I/II/III, *Compositio math.* **40** (1980), S. 35–67; **40** (1980), S. 69–78; **41** (1981), S. 23–30.
[19] Joseph A., Kostant's Problem, Goldie Rank and the Gelfand–Kirillov Conjecture, *Invent. math.* **56** (1980), S. 191–213.
[20] Joseph A., Dixmier's Problem for Verma and Principal Series Submodules, *J. London Math. Soc.* (2) **20** (1979), S. 193–204.
[21] Joseph A., W-Module Structure in the Primitive Spectrum of the Enveloping Algebra of a Semisimple Lie Algebra. In: *Non-Commutative Harmonic Analysis*, Lecture Notes in Mathematics **728**, Berlin–Heidelberg–New York, 1979, S. 116–135.
[22] Joseph A., Goldie Rank in the Enveloping Algebra of a Semisimple Lie Algebra I/II/III, *J. Algebra* **65** (1980), S. 269–283; **65** (1980), S. 284–306; **73** (1981), S. 295–326.
[23] Joseph A., Kostant's Problem and Goldie Rank. In: *Non-Commutative Harmonic Analysis and Lie Groups*, Lecture Notes in Mathematics **880**, Berlin–Heidelberg–New York, 1981, S. 249–266.
[24] Kazhdan D. A., Lusztig G., Representations of Coxeter Groups and Hecke Algebras, *Invent. math.* **53** (1979), S. 165–184.

[25] Springer T. A., Quelques applications de la cohomologie d'intersection, *Séminaire Bourbaki* 1981/82, exp. 589.
[26] Vogan D. A., A Generalized τ-Invariant for the Primitive Spectrum of a Semisimple Lie Algebra, *Math. Ann.* **242** (1979), S. 209–224.
[27] Vogan D. A., Ordering of the Primitive Spectrum of a Semisimple Lie Algebra, *Math. Ann.* **248** (1980), S. 195–203.
[28] Vogan D. A., Irreducible Characters of Semisimple Lie Groups II: The Kazhdan-Lusztig Conjectures, *Duke Math. J.* **46** (1979), S. 805–859.

MATHEMATISCHES INSTITUT DER UNIVERSITÄT
WEGELERSTRASSE 10, D-5300 BONN 1

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

A. JOSEPH

Primitive Ideals in Enveloping Algebras

1. Introduction

1.1. A basic problem in the theory of Lie algebras is to determine up to equivalence all irreducible representations of a given Lie algebra $\mathfrak{g}$. Here we shall always assume that $\mathfrak{g}$ is finite-dimensional over a field k which is algebraically closed and of characteristic zero. For finite-dimensional representations the problem quickly reduces to the semisimple case where the results are classical. For infinite-dimensional representations even the simplest non-commutative Lie algebras admit an enormously complicated representation theory. Now, the problem is equivalent to determining up to isomorphism all simple modules of the enveloping algebra $U(\mathfrak{g})$ of $\mathfrak{g}$. However, this latter viewpoint allows us to define a primitive ideal of $U(\mathfrak{g})$ to be the annihilator of a simple $U(\mathfrak{g})$ module and then to classify the set $\operatorname{Prim} U(\mathfrak{g})$ of all primitive ideals of $U(\mathfrak{g})$ which is better behaved. This procedure turns out to be rather like selecting out only the continuous representations of the corresponding Lie group and there is a remarkably similar but not quite precise correspondence between these two theories.

1.2. Dixmier [8] first drew serious attention to the study of $\operatorname{Prim} U(\mathfrak{g})$ and the early work was inspired by Kirillov's orbit method used in classifying unitary representations of real nilpotent (and subsequently solvable) Lie groups. Let G denote the algebraic adjoint group acting on the dual $\mathfrak{g}^*$ of $\mathfrak{g}$. Assume that $\mathfrak{g}$ is solvable. Then, by an appropriate application of Mackey's theory of induced representations and some further ideas from ring theory and algebraic geometry, one can construct a map $f \mapsto J(f)$ of $\mathfrak{g}^*$ onto $\operatorname{Prim} U(\mathfrak{g})$ which factors to a bijection of the orbit space $\mathfrak{g}^*/G$ onto $\operatorname{Prim} U(\mathfrak{g})$. For appropriate topologies this map is known to be continuous and even a homeomorphism for $\mathfrak{g}$ nilpotent [6]. The possible bicontinuity for $\mathfrak{g}$ solvable is still unresolved and involves delicate questions pertaining to inclusion relations between primitive ideals. The main part of this work was already completed by 1974 and has been fully described in [4, 8]. For more recent results see [36].

1.3. Already in 1978, Dixmier reported [7] on the considerable progress made in the classification of $\operatorname{Prim} U(\mathfrak{g})$ for arbitrary $\mathfrak{g}$. We are now in a position to describe a complete classification theory at least for $\mathfrak{g}$ algebraic. This consists of two parts: the semisimple case, and the reduction to the semisimple case. It is convenient to describe the two cases quite separately, as they follow a rather different philosophy. The former requires a quite new approach, whereas the latter still involves the method of induced representations; but is more subtle than the orbit method, and in particular, the relationship to $\mathfrak{g}^*/G$ becomes blurred. For example the intersection theorem of algebraic geometry carries over by bicontinuity to the nilpotent case, whereas it fails for $\mathfrak{g}$ semisimple (with respect to Gelfand–Kirillov dimension). Again, even disregarding the complication involved with the Goldie rank, an orbit in $\mathfrak{g}^*$ may give rise to more than one primitive ideal [20].

1.4. Returning to the question of describing all simple $U(\mathfrak{g})$ modules we let A_n denote the Weyl algebra of index n. It is the (non-commutative) algebra generated by the differential operators $\partial/\partial x_1, \partial/\partial x_2, \ldots, \partial/\partial x_n$ over the polynomial ring $S_n = k[x_1, x_2, \ldots, x_n]$. The representation theory of A_n is much simpler than that of an arbitrary enveloping algebra; but even this can be very complicated. Given $\mathfrak{g}$ nilpotent and $J \in \operatorname{Prim} U(\mathfrak{g})$, one has $U(\mathfrak{g})/J = A_n$ (where $2n = \dim Gf$ given $J = J(f)$). This reduces the study of simple $U(\mathfrak{g})$ modules with annihilator J to the study of simple A_n modules. In particular, S_n is a simple A_n module in an obvious fashion and leads to a simple $U(\mathfrak{g})$ module with annihilator J. Moreover, this module lifts in a suitable sense to a unitary representation of G and this construction gives rise to a bijection of $\operatorname{Prim} U(\mathfrak{g})'$ onto the unitary dual $\hat{G}$ of G. Though this is only a very simple example, it indicates the existence of a general principle, which is beginning to be more precisely formulated. Finally, for physicists, we remark that A_n is just the algebra of canonical commutation relations for n degrees of freedom and the above module was the one used by Heisenberg in describing the harmonic oscillator. There are many further correspondences and thus possibly room for application of the classification theory of $\operatorname{Prim} U(\mathfrak{g})$ in physics.

2. The semisimple case, preliminaries

2.1. The semisimple case is far too rich even to just describe all the results here. Consequently we shall concentrate on just one aspect of the classification theory, which was formulated in ([17], 7.4), a conjecture which has now been fully resolved. For a broader outlook we refer the reader to the review given in [23] and to Jantzen's forthcoming book [15].

2.2. Assume $\mathfrak{g}$ is semisimple. Let $\mathfrak{g} = \mathfrak{n}^+ \oplus \mathfrak{h} \oplus \mathfrak{n}^-$ be a triangular decomposition ([8], 1.10.14) with $R = R^+ \cup R^-$ being the corresponding decomposition of the root system R. Let $\varrho \in \mathfrak{h}^*$ denote the half-sum of the positive roots. For each $\alpha \in R$ set $\alpha^\vee := 2\alpha/(\alpha, \alpha)$ and let $s_\alpha \in \operatorname{Aut}\mathfrak{h}^*$ denote the reflection $\lambda \mapsto s_\alpha \lambda = \lambda - (\alpha^\vee, \lambda)\alpha$. The group W generated by the s_α, $\alpha \in R$, is called the *Weyl group* (for the pair $\mathfrak{g}$, $\mathfrak{h}$). It plays a fundamental role in the representation theory of $\mathfrak{g}$.

2.3. A *highest weight module* of highest weight $\lambda - \varrho \in \mathfrak{h}^*$ is any $U(\mathfrak{g})$ module generated by a vector e_λ (called a *highest weight vector*) satisfying $Xe_\lambda = 0$, $\forall X \in \mathfrak{n}^+$ and $He_\lambda = (\lambda - \varrho, H)e_\lambda$, $\forall H \in \mathfrak{h}$. For each $\lambda \in \mathfrak{h}^*$ there exists a unique universal highest weight module $M(\lambda)$ of highest weight $\lambda - \varrho$, and this admits a unique simple quotient $L(\lambda)$. We set $J(\lambda) = \operatorname{Ann} L(\lambda)$, which is a primitive ideal.

2.4. Let $Z(\mathfrak{g})$ denote the centre of $U(\mathfrak{g})$. According to Duflo ([10], II, Thm. 1) the map $\lambda \mapsto J(\lambda)$ of $\mathfrak{h}^*$ into $\operatorname{Prim} U(\mathfrak{g})$ is surjective. Composing this with the map $J \mapsto J \cap Z(\mathfrak{g})$ of $\operatorname{Prim} U(\mathfrak{g})$ into $\operatorname{Max} Z(\mathfrak{g})$ we get, by a result of Harish-Chandra ([8], 7.4) a surjection of $\mathfrak{h}^*$ onto $\operatorname{Max} Z(\mathfrak{g})$, which factors to a bijection $\mathfrak{h}^*/W \xrightarrow{\sim} \operatorname{Max} Z(\mathfrak{g})$. Consequently, $\operatorname{Prim} U(\mathfrak{g})$ is sandwiched between $\mathfrak{h}^*$ and $\mathfrak{h}^*/W$ and its classification reduces to describing the fibres of these maps.

2.5. Let $P(R) := \{\lambda \in \mathfrak{h}^* | (\alpha^\vee, \lambda) \in \mathbf{Z}, \forall \alpha \in R\}$ denote the lattice of integral weights. For expository purposes it is convenient to limit attention to this portion of $\mathfrak{h}^*$. In general similar results hold under appropriate modifications; but this is a subtle phenomenon and it is a good test of a proof if it works without this restriction. Set $P(R)^+ = \{\lambda \in P(R) | (\alpha, \lambda) \geqslant 0, \forall \alpha \in R^+\}$. It is a fundamental domain for the action of W on $P(R)$. Set $P(R)^{++} = \{\lambda \in P(R)^+ | (\alpha, \lambda) \neq 0, \forall \alpha \in R\}$, which forms the set of so-called *regular elements* of $P(R)^+$.

2.6. Fix $\lambda \in P(R)^+$ and let $\hat{\lambda}$ denote its image in $\mathfrak{h}^*/W$. Then $\operatorname{Ann}_{Z(\mathfrak{g})} M(w\lambda)$ is a maximal ideal $Z_{\hat{\lambda}}$ of $Z(\mathfrak{g})$ independent of the choice of $w \in W$. From this it is relatively easy to show that $M(w\lambda)$ has finite length with composition factors amongst the $L(y\lambda)$: $y \in W$. Let $b(w, y) := [M(w\lambda): L(y\lambda)]$ denote the number of times $L(y\lambda)$ occurs in the composition series for

$M(w\lambda)$. These numbers are independent of the choice of $\lambda \in P(R)^{++}$ (with some well-defined degeneration for $\lambda \in P(R)^{+}$). They play a fundamental role in the representation theory of $\mathfrak{g}$, and in particular in the description of $\operatorname{Prim} U(\mathfrak{g})$. Moreover, it was partly in an attempt to understand this role that led Kazhdan and Lusztig to formulate [27] their famous conjecture about these coefficients and which eventually led to their description in purely combinatorial terms. This work was reviewed by Jantzen in his talk at this congress [14]. It suffices to say that the $b(w, y)$ may be considered as known and, with respect to the Bruhat order, they form a matrix which is upper triangular with ones on the diagonal. It hence admits an inverse with integer coefficients which we denote by $a(w, y)$. Set

$$a(w) = \sum_{y \in W} a(w, y) y^{-1}, \quad w \in W.$$

These form a basis for the group ring $\boldsymbol{Z}W$.

2.7. Take $J \in \operatorname{Prim} U(\mathfrak{g})$. By Goldie's theorem, $U(\mathfrak{g})/J$ embeds in its ring of fractions, which is isomorphic to a matrix ring over a division algebra. The Goldie rank $\operatorname{rk}(U(\mathfrak{g})/J)$ of $U(\mathfrak{g})/J$ is defined to be the rank of this matrix ring. It is an important invariant of J. For example if $J = \operatorname{Ann} M$ for some simple finite-dimensional module M, then $\operatorname{rk} U(\mathfrak{g})/J = \dim M$ and yet $\operatorname{rk} U(\mathfrak{g})/J$ is finite even when $\dim M$ is infinite. In view of Duflo's parametrization of $\operatorname{Prim} U(\mathfrak{g})$ discussed in 2.4 it is natural to define for each $w \in W$ the function p_w on $P(R)^{+}$ by

$$p_w(\lambda) = \operatorname{rk}\big(U(\mathfrak{g})/J(w\lambda)\big).$$

We shall see that each p_w extends to a polynomial on $\mathfrak{h}^*$. These polynomials not only lead to a classification of $\operatorname{Prim} U(\mathfrak{g})$, but also to a remarkable connection with $\hat{W}$.

2.8. By the Poincaré–Birkhoff–Witt theorem, $U(\mathfrak{g})$ admits a filtration whose associated graded algebra identifies with the algebra of polynomial functions on $\mathfrak{g}^*$. Then for each $J \in \operatorname{Prim} U(\mathfrak{g})$ we can consider the variety $\mathfrak{V}(\operatorname{gr} J) \subset \mathfrak{g}^*$ of zeros of $\operatorname{gr} J$. We shall see that this variety is always irreducible and we describe it explicitly. (Irreducibility fails for $\mathfrak{g}$ solvable; but by a general result of Gabber [28] one always has equidimensionality). For this, let us identify $\mathfrak{g}^*$ with $\mathfrak{g}$ through the Killing form and call $X \in \mathfrak{g}$ *nilpotent* if $\operatorname{ad}_{\mathfrak{g}} X$ is a nilpotent endomorphism of $\mathfrak{g}$. The set $\mathfrak{N}$ of all nilpotent elements of $\mathfrak{g}$ is, of course, G-stable and forms a finite union of G orbits called *nilpotent orbits*. Through étale cohomology Springer [35] defined an injective map $\beta\colon \mathfrak{N}/G \to \hat{W}$. We shall see that this leads to a link with the correspondence described in 2.7, a totally unexpected and remarkable fact, which in its turn inspired a simpler version [21] of Springer's construction.

3. Semisimple case, results

3.1. Fix $w \in W$. The primitive ideals $J(w\lambda)$, $\lambda \in P(R)^+$, are pairwise distinct because their intersections with $Z(\mathfrak{g})$ are distinct maximal ideals (2.4). By the Borho–Jantzen translation principle ([5], 2.12) the inclusion relations in the fibre $\mathfrak{X}_\lambda := \{J(w\lambda)\colon w \in W\}$ over Z_λ are independent of the choice of $\lambda \in P(R)^{++}$ (with some well-defined degeneration for $\lambda \in P(R)^+$). We might therefore guess that these relations can be described purely combinatorially in terms of W. Indeed, for any subset $S \subset \mathbf{Z}W$ let $[S]$ denote the subset of $\{a(y)\colon y \in W\}$ which occur with non-zero coefficient in the expansion of some $s \in S$ as a linear combination of the $a(y)\colon y \in W$.

PROPOSITION. *Assume $\lambda \in P(R)^{++}$. For each pair $w,y \in W$ one has*

$$J(w\lambda) \supset J(y\lambda) \Leftrightarrow a(w) \in [a(y)W].$$

This question was formulated and the implication $\Leftarrow$ was established in [18]. The conjectured reverse implication was established by Vogan [37]. It shows that the $a(w)$ completely determine the inclusion relations between primitive ideals, which in view of Kazhdan–Lusztig's theorem can be described purely combinatorially. Lusztig [29] used this result to establish an isomorphism between a Hecke algebra derived from W and the group ring $\mathbf{Q}(t)W$. This was a purely combinatorial question which needed Prim $U(\mathfrak{g})$ for its solution!

3.2. Fix $w \in W$. Since ϱ is regular, one easily checks that $a(w)\varrho^m$ cannot vanish for all $m \in \mathbf{N}$. Let n denote the least non-negative integer with this property. Define p_w as in 2.7. Then ([19], II, 5.1, 5.5)

THEOREM. *For each $w,y \in W$, one has*
(i) $p_w = a(w)\varrho^n$, *up to a scalar.*
(ii) *Suppose $\lambda \in P(R)^{++}$. Then* $p_w(\lambda) = p_y(\lambda) \Leftrightarrow J(w\lambda) = J(y\lambda)$.

The determination of the scalars in (i) has not yet been fully completed. It is a delicate question which involves in part the construction of sufficiently many completely prime, primitive ideals (i.e. for which the quotient algebra has Goldie rank one). Apart from this, we see that the $a(w)$ determine the Goldie ranks of the primitive quotients, which by (ii) separate the elements of $\mathfrak{X}_{\hat{\lambda}}$. This separation is better than that given in 3.1, which is highly implicit; for example, it allows one to calculate card $\mathfrak{X}_{\hat{\lambda}}$ (see 3.3). One can have $p_w(\lambda) = 0$ if $\lambda \in P(R)^+$ and this exactly describes the degeneration of the fibre $\mathfrak{X}_{\hat{\lambda}}$ alluded to in 3.1.

3.3. A deep consequence of 3.1 and 3.2 is that QWp_w is a simple W module P_τ say of type $\tau \in \hat{W}$. Set $(\mathfrak{X}_\lambda)_\tau = \{J(w\lambda) | p_w \in P_\tau\}$. Note that this is actually a partition of W into so-called *double cells*. Then ([19], II, 5.5)

THEOREM. *The distinct* p_y: $J(y\lambda) \in (\mathfrak{X}_{\hat{\lambda}})_\tau$ *form a basis for* P_τ. *In particular if* $\lambda \in P(R)^{++}$ *then* $\operatorname{card}(\mathfrak{X}_{\hat{\lambda}})_\tau = \deg \tau$.

This result may be restated by saying that basis vectors for certain irreducible representations of W classify $\mathfrak{X}_{\hat{\lambda}}$ and hence $\operatorname{Prim} U(\mathfrak{g})$. These bases are implicitly determined by 3.2(i); but their explicit description is still an interesting open question. Except for type A_n, not all irreducible representations of W occur as a P_τ. Those that do occur are again implicitly determined by 3.2. From this and by some case by case analysis, Barbasch and Vogan [1, 2] showed that these representations are just the special representations in the sense of Lusztig, whose definition is given in terms of the classification of irreducible representations of finite Chevalley groups. This remarkable coincidence is not yet fully understood.

3.4. Recall the definition of the map β (2.8).

THEOREM. *Take* $\tau \in \hat{W}$ *occurring as a* P_τ. *Then* $\tau \in \operatorname{Im}\beta$ *and for each* $J \in (\mathfrak{X}_{\hat{\lambda}})_\tau$, $\mathfrak{V}(\operatorname{gr} J)$ *is the Zariski closure of* $\beta^{-1}(\tau)$.

In particular, $\mathfrak{V}(\operatorname{gr} J)$ is independent of $\lambda \in P(R)^+$ and $J \in (\mathfrak{X}_\lambda)_\tau$. This was proved in ([16], 2.6). From this, by the result for induced ideals and some case by case manipulations, the above theorem was established in the present so-called integral case by Borho and Brylinski [3]. A general proof was given in [24]. It was based on Gabber's equidimensionality theorem (2.8) and a reinterpretation of Springer's construction formulated in [21] and established by Hotta [13]. The proof uses a deep separation theorem of Gabber ([25], Sect. 7). It is worth noting that not all nilpotent orbits occur in the integral fibres $\mathfrak{X}_{\hat{\lambda}}$, $\lambda \in P(R)^+$. However if one takes the union over all possible fibres $\mathfrak{X}_{\hat{\lambda}}$, $\lambda \in \mathfrak{h}^*$, then one expects to find all the nilpotent orbits. This does not seem to have yet been checked but in any case has no elegant proof.

3.5. A more delicate question than the above is to be able to associate a completely prime primitive ideal to each nilpotent orbit. This is an open problem; but it is known that one can sometimes have two ([20], Sect. 4). A further open problem is to determine the composition factors of primitive quotients considered as $U(\mathfrak{g})-U(\mathfrak{g})$ bimodules. One expects the answer to be determined by the $a(w)$, see [22].

3.6. Returning to the question raised in 1.4 we remark that the Conze embeddnig [26] of $U(\mathfrak{g})/\mathrm{Ann}M(\lambda)$ into A_n with $n = \mathrm{card}\,R^+$ makes A_n flat left $U(\mathfrak{g})/\mathrm{Ann}M(\lambda)$ module for λ regular. Combined with work of Hodges and Smith this eventually gives $\mathrm{gl\,dim}\,U(\mathfrak{g})/\mathrm{Ann}M(\lambda) \leqslant \mathrm{card}\,R$ for λ regular. Yet $\mathrm{gl\,dim}\,U(\mathfrak{g})/\mathrm{Ann}M(\lambda) = \infty$ for λ non-regular.

4. Reduction to the semisimple case

4.1. It is assumed that $\mathfrak{g} = \mathrm{Lie}(G)$ where G is a connected linear algebraic group. This is a slight restriction, which can probably be overcome without much extra difficulty. Note that G acts by automorphisms in $U(\mathfrak{g})$ and because G is connected, any ideal of $S(\mathfrak{g})$ or any two-sided ideal of $U(\mathfrak{g})$ is G stable ([4], 12.3).

4.2. Let $\mathfrak{u}$ be an ideal of $\mathfrak{g}$. If $J \in \mathrm{Prim}\,U(\mathfrak{g})$, then $J \cap U(\mathfrak{u})$ is prime ([8], 3.3.4) but it need not be primitive. It is therefore useful to be able to recognize, which prime ideals are primitive. This is provided by the following result established by Dixmier ([9], Thm. C) for $k = \boldsymbol{C}$ and by Moeglin ([30], Sec. 4) in general.

THEOREM. *The following two conditions are equivalent*
(i) $J \in \mathrm{Prim}\,U(\mathfrak{g})$.
(ii) $J \in \mathrm{Spec}\,U(\mathfrak{g})$ *and* $(\mathrm{Fract}\,U(\mathfrak{g})/J)^G$ *is reduced to scalars.*

4.3. Let $\mathfrak{u}$ be an ideal of $\mathfrak{g}$ and take $J \in \mathrm{Prim}\ U(\mathfrak{g})$. By 4.2 it easily follows that $(\mathrm{Fract}\,U(\mathfrak{u})/U(\mathfrak{u})\cap J)^G = k$. In the commutative case this would, imply the existence of a unique dense G orbit in the subvariety of $\mathfrak{u}^*$ of zeros of $J\cap U(\mathfrak{u})$. The non-commutative case was first studied by Dixmier [9] and later completely analysed by Moeglin and Rentschler [32]. One has the following

THEOREM. *Take* $J \in \mathrm{Prim}\,U(\mathfrak{g})$.
(i) ([32], (2)). *There exists* $K \in \mathrm{Prim}\,U(\mathfrak{u})$ *such that* $J\cap U(\mathfrak{u}) = \bigcap_{g\in G} gK$.

(ii) ([32], (1) bis). *Any two primitive ideals of* $U(\mathfrak{u})$ *satisfying* (i) *are conjugate under* G.

4.4. Preserve the notation and hypotheses of 4.3. Set $B = \{b \in G | bK = K\}$. This is an algebraic subgroup of G which need not be connected. Set $\mathfrak{b} = \mathrm{Lie}(B) = \{X \in \mathfrak{g} | [X, K] \subset K\}$. If I is a two-sided ideal of $U(\mathfrak{b})$ we set $\mathrm{Ind}(I, \mathfrak{b}\uparrow\mathfrak{g}) = \bigcap_{g\in G} g\big(U(\mathfrak{g})I\big)$. Again by Moeglin and Rentschler [32, 33] we have the following

THEOREM. (i) ([32], (3)) *There exists* $I \in \mathrm{Prim}\, U(\mathfrak{b})$ *such that* $I \cap U(\mathfrak{u}) = K$ *and* $J = \mathrm{Ind}(I, \mathfrak{b}\uparrow\mathfrak{g})$.

(ii) ([33]) *Any two primitive ideals of* $U(\mathfrak{b})$ *satisfying* (i) *are conjugate under* B. *In particular there are only finitely many of them.*

4.5. From now on we take $\mathfrak{u}$ to be the unipotent radical of $\mathfrak{g}$. Since $K \in \mathrm{Prim}\, U(\mathfrak{u})$, we can write $K = J(u)$ for some $u \in \mathfrak{u}^*$ (see 1.2). As a consequence of 4.3 (ii), the G orbit of u in $\mathfrak{u}^*_-$ is completely determined by J. Set $\mathfrak{h} = \mathrm{Stab}_{\mathfrak{g}}(u) := \{X \in \mathfrak{g} | Xu = 0\} = \{X \in \mathfrak{g} | u([X, \mathfrak{u}]) = 0\}$. One has $\mathfrak{b} = \mathfrak{h} + \mathfrak{u}$, and obviously I contains the two-sided ideal $U(\mathfrak{b})J(u)$. Set

$$L_u = \sum_{X\in\mathfrak{h}\cap\mathfrak{u}} U(\mathfrak{h})\big(X - u(X)\big)$$

which is again a two-sided ideal of $U(\mathfrak{h})$. According to Duflo ([8], 10.1.4,) we have

PROPOSITION. *There is a canonical algebra isomorphism*

$$D_u \text{ of } U(\mathfrak{b})/U(\mathfrak{b})J(u) \text{ onto } U(\mathfrak{h})/L_u \otimes_k U(\mathfrak{u})/J(u).$$

4.6. Apart from a cohomology obstruction (here it is needed that $u([\mathfrak{h}, \mathfrak{u}]) = 0$), the existence statement in 4.5 results from the fact that $\mathfrak{h}$ acts by derivations on $U(\mathfrak{u})/J(u)$, which, (1.4) being a Weyl algebra A_n, admits only inner derivations ([8], 4.6.8). Again A_n is central simple, and so D_u sets up a bijection (which we shall also denote by D_u) from the set of two-sided ideals of $U(\mathfrak{b})$ containing $U(\mathfrak{b})J(u)$ onto the set $\mathrm{Prim}_u U(\mathfrak{h})$ of two-sided ideals of $U(\mathfrak{h})$ containing L_u. This takes primes to primes and so, by 4.1, primitives to primitives. Furthermore, Blattner's criterion ([8], 5.3.6) gives

THEOREM. *If* $J_1 \in \mathrm{Prim}_u U(\mathfrak{h})$, *then* $J_0 := \mathrm{Ind}\big(D_u^{-1}(J_1), \mathfrak{b}\uparrow\mathfrak{g}\big) \in \mathrm{Prim}\; U(\mathfrak{g})$.

4.7. In the above manner the classification of Prim $U(\mathfrak{g})$ is reduced to that of Prim $U(\mathfrak{h})$ for some, usually smaller, subalgebras $\mathfrak{h}$. Indeed, for a given $J \in \operatorname{Prim} U(\mathfrak{g})$ this process gives a sequence $\mathfrak{u}_1 = \mathfrak{u}, \mathfrak{u}_2, \ldots,$ of unipotent Lie algebras, of linear forms $u_1 = u, u_2, \ldots, u_i \in \mathfrak{u}_i^*$, of algebraic groups $B_1 = B, B_2, \ldots,$ of algebraic groups $H_0 = G, H_1, \ldots,$ of primitive ideals $J_0 = J, J_1, \ldots, J_i \in \operatorname{Prim} U(\mathfrak{h}_i)$ and of primitive ideals $I_1 = I, I_2, \ldots$ $\ldots, I_i \in \operatorname{Prim} U(\mathfrak{b}_i)$, where $B_i = \operatorname{Stab}_{B_{i-1}}(J(u_i))$, $\mathfrak{b}_i = \operatorname{Lie}(B_i)$, H_i is the identity component of $\operatorname{Stab}_{H_{i-1}}(u_i)$, $\mathfrak{h}_i = \operatorname{Lie}(H_i)$, $\mathfrak{u}_{i+1}$ is the unipotent radical of $\mathfrak{h}_i$, $u_{i+1} \in \mathfrak{u}_{i+1}^*$ is in the unique H_i orbit determined by J_i (recall 4.5), I_i is in the unique B_i-orbit (4.4) determined by the pair $(J_{i-1}, J(u_i))$ and $J_i = D_{u_i}(I_i)$. The process stops when $\mathfrak{h}_t \cap \mathfrak{u}_t$ is complemented in $\mathfrak{h}_t$ by a reductive Lie algebra $\mathfrak{r}$ (Levi factor). Set $\mathfrak{n} = \sum \mathfrak{u}_i$, $\mathfrak{l} = \mathfrak{h}_t$, $\mathfrak{p} = \mathfrak{l} + \mathfrak{n}$. Observe that $\mathfrak{n}$ is the unipotent radical of $\mathfrak{p}$. From the above construction it follows that there exists $n \in \mathfrak{n}^*$ such that $n|_{\mathfrak{u}_i} = u_i$. Then the above process may be collapsed to give $J = \operatorname{Ind}(D_n^{-1}(\tilde{J}_t), \mathfrak{p}\uparrow\mathfrak{g})$ ([12], Chap. IV, Sec. 9). Here $\tilde{J}_t$ differs very slightly from J_t (just with respect to the centre of $\mathfrak{r}$). This difficulty may be overcome by replacing Ind everywhere by $\operatorname{Ind}^{\sim}$ (see [11] for definition and subtleties involved).

4.8. We now describe the parametrization of Prim $U(\mathfrak{g})$ which results from the above reduction. Take $g \in \mathfrak{g}^*$ extending $n \in \mathfrak{n}^*$. Observe that we can recover the sequence $\mathfrak{h}_0 = \mathfrak{g}, \mathfrak{h}_1, \mathfrak{h}_2, \ldots,$ by setting $\mathfrak{h}_i = \operatorname{Stab}_{\mathfrak{h}_{i-1}}(g|_{\mathfrak{u}_i})$ where $\mathfrak{u}_i$ is the unipotent radical of $\mathfrak{h}_{i-1}$. Then, with $\mathfrak{p}$, $\mathfrak{n}$, $\mathfrak{l}$, $\mathfrak{r}$ as above we note that $g([\mathfrak{l}, \mathfrak{n}]) = 0$. Furthermore, one checks, with respect to the form $(X, Y) \mapsto g([X, Y])$ defined by g, that $\mathfrak{p}^\perp \subset \mathfrak{h}_1$; continuing inductively we obtain $\mathfrak{p}^\perp \subset \mathfrak{l}$; so, in particular, $\mathfrak{p}$ is co-isotropic. Let $\mathfrak{g}(g)$ denote the kernel of this form and define $\mathfrak{h}(h)$ $(h := g|_{\mathfrak{h}})$ and $\mathfrak{u}(u)$ similarly. An elementary calculation gives $\mathfrak{h}(h) = \mathfrak{g}(g) + \mathfrak{u}(u)$, which applied inductively gives $\mathfrak{l}(l) \subset \mathfrak{g}(g) + \mathfrak{n}$ (where $l = g|_{\mathfrak{l}}$). Finally, assume that we have taken the extension g of n to satisfy $g(\mathfrak{r}) = 0$. One easily checks that $\mathfrak{r} \subset \mathfrak{l}(l)$ and so $\mathfrak{p} = \mathfrak{g}(g) + \mathfrak{n}$.

4.9. Given $g \in \mathfrak{g}^*$, we define (following Duflo [12], I. 8) a subalgebra $\mathfrak{p} \subset \mathfrak{g}$ to be of strongly unipotent type with respect to g if $\mathfrak{p}$ is algebraic, co-isotropic with respect to g and $\mathfrak{p} = \mathfrak{g}(g) + \mathfrak{n}$ where $\mathfrak{n}$ is the unipotent radical of $\mathfrak{p}$. We define $g \in \mathfrak{g}^*$ to be of *Duflo type* if

(i) There exists a Levi factor $\mathfrak{r}$ of $\mathfrak{g}(g)$ satisfying $g(\mathfrak{r}) = 0$. (If this holds for one factor, it holds for all of them since they are conjugated under $G(f) := \operatorname{Stab}_G(g)$.)

(ii) There exists a subalgebra $\mathfrak{p}$ of $\mathfrak{g}$ of strongly unipotent type with respect to g.

We have seen that g in 4.8 is of Duflo type. Again if $g \in \mathfrak{g}^*$ is of Duflo type then $g|_{\mathfrak{h}}$ is of Duflo type ([12], Lemme 17). This is used to show that if $g \in \mathfrak{g}^*$ is of Duflo type then the $\mathfrak{p}$ constructed canonically in 4.8 is of strongly unipotent type with respect to g. Finally, Duflo ([12], Prop. 26 (i)) showed that if $g \in \mathfrak{g}^*$ is of Duflo type and $\mathfrak{p}$ is of strongly unipotent type with respect to g, then g is determined up to conjugation under G by its restriction to $\mathfrak{p}$.

4.10. Let $\mathfrak{g}_D^*$ denote the subset of linear forms on $\mathfrak{g}$ of Duflo type. Let $\mathfrak{p}, \mathfrak{n}, \mathfrak{l}, n$ be as defined in 4.8. Let $\mathfrak{r}_g$ be a Levi factor of $\mathfrak{g}(g)$ (we remind that they are all conjugate under $G(g)$); it is also a Levi factor of $\mathfrak{l}$. Given $P \in \operatorname{Prim} U(\mathfrak{r})$, let Q be the unique primitive ideal of $U(\mathfrak{l})$ containing

$$L_n = \sum_{X \in \mathfrak{l} \cap \mathfrak{n}} U(\mathfrak{l})\big(X - n(X)\big)$$

such that $Q \cap U(\mathfrak{r}) = P$. Set $J(g, P) = \operatorname{Ind}^{\sim}\big(D_n^{-1}(Q), \mathfrak{p}\uparrow\mathfrak{g}\big)$. Then

THEOREM. (i) *Every* $J \in \operatorname{Prim} U(\mathfrak{g})$ *takes the form* $J = J(g,P)$ *for some* $g \in \mathfrak{g}_D^*$, $P \in \operatorname{Prim} U(\mathfrak{r}_g)$.

(ii) *Given* $J(g, P) = J(g', P')$, *there exists* $s \in G$ *such that* $sg = g'$, $s\mathfrak{r}_g = \mathfrak{r}_{g'}$, $sP = P'$.

This means that the map $(g, P) \mapsto J(g, P)$ factors to a bijection of $\coprod_{g \in \mathfrak{g}_D^*} \{P \in \operatorname{Prim} U(\mathfrak{r}_g)\}/G$ onto $\operatorname{Prim} U(\mathfrak{g})$. Surjectivity is due to Duflo ([12], IV, Thm. 7) and injectivity to Moeglin and Rentschler [33]. For injectivity the crucial part was 4.4 (ii).

4.11. Retain the notation of 4.10. It is easy to see that $\operatorname{rk}\big(U(\mathfrak{r}_g)/P\big) \geqslant \operatorname{rk}\big(U(\mathfrak{g})/J(g, P)\big)$, $\forall g \in \mathfrak{g}_D^*$, $P \in \operatorname{Prim} U(\mathfrak{r}_g)$; but it is not quite obvious if equality holds. Again for each $g \in \mathfrak{g}_D^*$ the map $P \mapsto J(g, P)$ factorizes to an order isomorphism of $\big(G(g)\operatorname{Prim} U(\mathfrak{r}_g)\big)/G(g)$ onto its image in $\operatorname{Prim} U(\mathfrak{g})$. Here the order is defined by inclusion. It is not yet clear what might be the order relations with respect to the different $g \in \mathfrak{g}_D^*$, though one should expect these to be expressible, at least partly, in terms of the inclusion relations of the Zariski closures of the orbits Gg. For $\mathfrak{g}$ solvable such a result would be equivalent to the bicontinuity of the map $\mathfrak{g}^*/G \xrightarrow{\sim} \operatorname{Prim} U(\mathfrak{g})$ (recall 1.2) and this is still an open question.

Section 4 is based on a talk of Rentschler given at Oberwolfach [34] in January 1983. I should like to thank Rentschler for explaining the results he obtained in collaboration with Moeglin [32, 33] and the work of Duflo [12]. Rentschler has recently reported that the restriction on g being algebraic can be removed and has described the modifications that ensue.

References

[1] Barbasch D. and Vogan D., Primitive Ideals and Orbital Integrals in Complex Classical Groups, *Math. Ann.* **259** (1982), pp. 153–199.

[2] Barbasch D. and Vogan D., Primitive Ideals and Orbital Integrals in Complex Exceptional Groups, *J. Algebra* **80** (1983), pp. 350–382.

[3] Borho W. and Brylinski J.-L., Differential Operators on Homogeneous Spaces. I, *Invent. Math.* **69** (1982), pp. 437–476.

[4] Borho W., Gabriel P., and Rentschler R., *Primideale in Einhüllenden auflösbarer Lie Algebren,* LN 357, Springer-Verlag, Berlin 1973.

[5] Borho W. and Jantzen J.-C., Über primitive Ideale in der Einhüllenden einer halbeinfachen Lie-Algebra, *Invent. Math.* **39** (1977), pp. 1–53.

[6] Conze N., Espace des idéaux primitifs de l'algèbre enveloppante d'une algèbre de Lie nilpotente, *J. Algebra* **34** (1975), pp. 444–460.

[7] Dixmier J., Algèbres enveloppantes. In: *Proceedings of the International Congress of Mathematicians,* Helsinki, 1978, pp. 693–698.

[8] Dixmier J., *Algèbres enveloppantes,* Gauthier-Villars, Paris 1974.

[9] Dixmier J., Idéaux primitifs dans les algèbres enveloppantes, *J. Algebra* **48** (1977), pp. 86–112.

[10] Duflo M., Sur la classification des idéaux primitifs dans l'algèbre enveloppante d'une algèbre de Lie semi-simple, *Ann. Math.* **105** (1977), pp. 107–120.

[11] Duflo M., Sur les idéaux induits dans les algèbres enveloppantes, *Invent. Math.* **67** (1982), pp. 385–392.

[12] Duflo M., Théorie de Mackey pour les groupes de Lie algébriques, *Acta Mathematica* (to appear) .

[13] Hotta R., *On Joseph's Construction of Weyl Group Representations,* preprint, Tohoku, 1982.

[14] Jantzen J.-C., Einhüllende Algebren halbeinfacher Lie-Algebren, *Proceedings of the International Congress of Mathematicians, Warsaw, 1983.*

[15] Jantzen J.-C., Einhüllende Algebren halbeinfacher Lie-Algebren, *Ergebnissen der Mathematik,* Springer-Verlag, 1983.

[16] Joseph A., Towards the Jantzen Conjecture. III, *Compos. Math.* **41** (1981), pp. 23–30.

[17] Joseph A., Kostant's Problem, Goldie Rank and the Gelfand–Kirillov Conjecture, *Invent. Math.* **56** (1980), pp. 191–213.

[18] Joseph A., *W- Module Structure in the Enveloping Algebra of a Semisimple Lie Algebra,* LN 728, Springer-Verlag, Berlin 1979.

[19] Joseph A., Goldie Rank in the Enveloping Algebra of a Semisimple Lie Algebra, I/II/III, *J. Algebra* **65** (1980), pp. 269–283; pp. 284–306; **73** (1981), pp. 295–326.

[20] Joseph A., *Kostant's Problem and Goldie Rank,* LN 880, Springer-Verlag, Berlin 1981.

[21] Joseph A., On the Variety of a Highest Weight Module, *J. Algebra* (to appear).

[22] Joseph A., Completion Functors in the 0 Category. In: *Non-Commutative Harmonic Analysis and Lie Groups,* Springer Lecture Notes 1020, 1983.

[23] Joseph A., On the Classification of Primitive Ideals in the Enveloping Algebra of a Semisimple Lie Algebra: *Lie Group Representations I,* Springer Lecture Notes 1024, 1983.

[24] Joseph A., On the Associated Variety of a Promitive Ideal, to appear in *J. of Algebra.*

[25] Joseph A., *Application de la theorie des anneaux aux algèbres enveloppantes,* notes de 3eme cycle, Paris, 1981.

[26] Joseph A. and Stafford T. J., to appear.

[27] Kazhdan D. A. and Lusztig G., Representations of Coxeter Groups and Hecke Algebras, *Invent. Math.* **53** (1979), pp. 165–184.

[28] Levasseur T., *Equidimensionalité de la variété caractéristique,* preprint, Paris, 1983.

[29] Lusztig G., On a Theorem of Benson and Curtis, *J. of Algebra* **71** (1981), pp. 490–498.

[30] Moeglin C., Idéaux primitifs dans les algèbres enveloppantes, *J. Math. pures et appl.* **59** (1980), pp. 265–336.

[31] Moeglin C., Idéaux bilatères des algèbres enveloppantes, *Bull. Soc. Math. France* **108** (1980), pp. 143–186.

[32] Moeglin C. and Rentschler R., Orbites d'un groupe algèbrique dans l'espace des idéaux rationnels d'une algèbre enveloppantes, *Bull. Soc. Math. France* **109** (1981), pp. 403–426.

[33] Moeglin C. and Rentschler R., *Sur la classification des idéaux primitifs des algèbres enveloppantes,* preprint, Paris, 1983.

[34] Rentschler R., *Report on Primitive Ideals in Enveloping Algebras of Lie Algebras,* talk, Oberwolfach, 1983.

[35] Springer T. A., A Construction of Representations of Weyl Groups, *Invent. Math.* **44** (1978), pp. 279–293.

[36] Tauvel P., Sur la bicontinuité de l'application de Dixmier pour les algèbres de Lie résolubles, *C. R. Acad. Sci. Paris* **294** (1982), pp. 685–687.

[37] Vogan D. A., Ordering of the Primitive Spectrum of a Semisimple Lie Algebra, *Math. Ann.* **248** (1980), pp. 195–203.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

A. YU. OL'ŠANSKIĬ

On a Geometric Method in the Combinatorial Group Theory

The purpose of this report is to draw attention to a fruitful geometric method of the infinite group theory which has led the author to solving a number of group-theoretic problems. Among these are: the problem of O. Yu. Schmidt about the existence of an infinite non-abelian group in which all proper subgroups are finite [8]–[10], the problem on the existence of a non-amenable group without free subgroups [12] attributed to John von Neumann [2] and the construction of the 'Tarski-monster' [14]. We will also mention a new relatively short proof of a theorem due to Novikov–Adyan [13] and some other results.

The progress of topological investigations was one of the main stimuli for studying groups given by generators and defining relations. First connections of questions of the combinatorial topology with group-combinatorial ones were established by H. Poincaré, M. Dehn and other mathematicians on the eve of this century. First problems of the combinatorial group theory originate in geometry and topology.

Traditionally, many algebraic systems arose in connection with classification of geometric objects. But the situation when topological ideas help to answer a question in algebra is rare. Therefore, the simple but important observation of van Kampen [18] has perhaps been overlooked by many authors. The essence of the van Kampen lemma is a clear geometric interpretation of deducibility consequences of defining relations. Before we pass to precise definitions we illustrate a simple case; for example, the equation

$$a^2ba^{-4}b^2ab^{-1}a^{-1}b^{-1}a^{-1}b^{-1} = 1$$

follows from relations $a^3 = 1$ and $aba^{-1}b^{-1} = 1$, i.e., $ab = ba$ (Fig. 1).

We read one of the defining relations by going around any region and we read a consequence by going around the boundary of the map (passing the inverse edge gives the inverse letter).

The universal character of the van Kampen lemma makes the geometric approach to studying group presentation very natural. This approach yields an attraction of elementary facts in the combinatorial topology (first of all, the Euler formula and the Jordan lemma) for solving some problems arising in algebra. In the middle of the sixties the van Kampen lemma was discovered and successfully applied by a number of mathematicians to the Small Cancellation Theory, the central ideas of which are treated in the book [5] of R. Lyndon and P. Schupp.

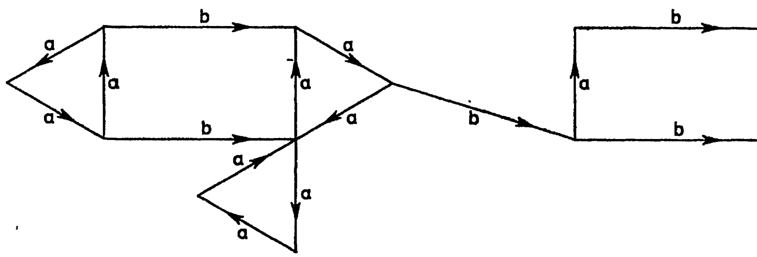


Fig. 1

Following [5], we give some definitions. Let $\boldsymbol{R}^2$ denote the Euclidean plane. For a subset $S \subseteq \boldsymbol{R}^2$ we denote the boundary of S by ∂S, and the closure of S by $\bar{S}$. A *map* M is a finite collection of vertices (points of $\boldsymbol{R}^2$), edges (bounded subsets homeomorphic to the open unit interval) and cells (bounded subsets homeomorphic to the open unit disk) that are pairwise disjoint and satisfy the following conditions: (i) if e is an edge of M, then there exist vertices a and b such that $\bar{e} = e \cup \{a\} \cup \{b\}$, (ii) the boundary $\partial \Pi$ of each cell Π of M is $\bar{e}_1 \cup \bar{e}_2 \cup \ldots \cup \bar{e}_n$ for some edges $e_1, \ldots, e_n$ of M. Edges of maps will be viewed as having two possible orientations.

Let $\{a_1, a_2, \ldots\}$ be an alphabet and let a group G be defined by its presentation:

$$\langle a_1, a_2, \ldots | \; R_1 = 1, \; R_2 = 1, \ldots \rangle, \tag{1}$$

where $\{R_i\}_{i=1}^{\infty} = \mathscr{R}$ is the set of defining words of G. By a *diagram* Δ over $\mathscr{R}$ we mean an oriented, connected, simply connected map M and a function φ such that (i) the function φ assigns to each oriented edge e of Δ one of the letters $a_1^{\pm 1}, a_2^{\pm 1}, \ldots$ (the *label* of e) and $\varphi(e^{-1}) = \varphi(e)^{-1}$, (ii) if $e_1 \ldots e_n$ is the boundary path of some cell Π of Δ, then $\varphi(e_1) \ldots \varphi(e_n) \overset{\circ}{=} R$ is a cyclic shift of some $R_i^{\pm 1}$, $i = 1, 2, \ldots$ ($\overset{\circ}{=}$ denotes graphical equality).

The statement of the van Kampen lemma is almost obvious: a relation $W = 1$ is a consequence of (1), iff there exist a diagram Δ over (1) with the boundary path $e_1 \dots e_m$ of Δ such that $\varphi(e_1)\varphi(e_2)\dots\varphi(e_m) \overset{\circ}{=} W$ [5]. Moreover, one may assume that Δ is reduced, i.e., there is no pair of cells Π_1, Π_2 with a common edge $e \subseteq \partial\Pi_1 \cap \partial\Pi_2$ such that for boundary paths ep_1, $e^{-1}p_2^{-1}$ of Π_1, Π_2 the equation $\varphi(p_1) = \varphi(p_2)$ holds in the free group.

It is convenient to define $\mathscr{R}^*$ as a set of cyclic shifts of words $R^{\pm 1} \in \mathscr{R}$. A common initial segment of two different words of $\mathscr{R}^*$ is called a *piece*. The most popular *condition* $C'(\lambda)$ of the Small Cancellation Theory says that if U is a piece and $R \overset{\circ}{=} UV \in \mathscr{R}^*$, then $|U| < \lambda |R|$.

Throughout this paper, a maximal subpath p of the boundary path of a cell Π such that p^{-1} is a subpath of some cell Π' or p^{-1} is a subpath of the boundary path of Δ will be called an *arc* of Π. The number of arcs of Π will be called a *degree* of Π. The condition $C'(\lambda)$ means that the degree of every interior cell of a van Kampen diagram is greater than λ^{-1}. If λ is small, then Euler's formula proves the existence of a cell Π with a 'long' external arc p (for instance, $|p| > \frac{1}{2}|\partial\Pi|$). It is clear that we have a basis for Dehn's equality algorithm: a non-trivial cyclic reduced word W equals 1 in G iff W contains a cyclic subword U of some $R \in \mathscr{R}^*$ and $|U| > \frac{1}{2}|R|$. M. Dehn applied the algorithm for the fundamental groups of compact Riemann surfaces of genus $g > 1$ (then the condition $C'\big(1/(4g-1)\big)$ holds).

Many important and typical results of Small Cancellation Theory are exibited in [5]. Contributions to this theory are made by V. A. Tartakovskij, M. D. Greendlinger, J. L. Britton, H. Schiek, R. Lyndon, C. M. Weinbaum, P. E. Schupp, C. Lipschutz, A. I. Gol'berg and some others (see [5]).

Many important groups, however, have no presentation with $C'(\lambda)$ for small λ, or similar conditions. Firstly, the groups with small cancellation conditions contain free non-cyclic subgroups and many normal subgroups. Secondly, we have the following example. In order to construct a non-trivial finitely generated periodical group it is natural to try to define the set $\mathscr{R}$ of defining words as follows. Let $R_1 \overset{\circ}{=} a_1^{n_1}$, where n_1 is a sufficiently large number. Having defined $R_1, \dots, R_{i-1} \in \mathscr{R}$, let A_i be the shortest word of infinite order in $G_{i-1} = \langle a_1, \dots, a_m \mid R_1 = 1, \dots, R_{i-1} = 1\rangle$. We put $R_i \overset{\circ}{=} A_i^{n_i}$. It is clear that the presentation gives a periodical group if $\mathscr{R} = \bigcup_{i=1}^{\infty} \{R_i\}$. But it is impossible to use any $C'(\lambda)$ condition for proving the group being infinite. Indeed, if $R_j \overset{\circ}{=} UV$ and $|U| \approx \frac{1}{2}|R_j|$, then it is possible that for some $i \gg j$, $A_i \overset{\circ}{=} UV'$, i.e., U is a piece.

We develope the above geometric method for constructing groups with completely new properties. Thus, we obtain answers to the questions whose appearance was very natural in the period, when the group theory was passing from finite to infinite objects. Their solutions required the introduction of new concepts. We formulate at first the theorems concerning the existence of groups with a very simple subgroup structure.

A *Noetherian group* is a group with no infinite ascending chain of subgroups, i.e., each of its subgroup is finitely generated. The following question of Baer is well known: Does an arbitrary Noetherian group G possess a finite series $\{1\} = G_0 \lhd G_1 \lhd \ldots \lhd G_n = G$, where each factor G_k/G_{k-1} is a finite or a cyclic group? (The converse is obviously true.)

THEOREM 1 ([8],[9]). *There exists an infinite simple Noetherian group G. Moreover,* (i) *G is effectively defined and there exist equality and conjugacy algorithms for G,* (ii) *each proper subgroup of G is infinite cyclic,* (iii) *extraction of roots in G is unique: $X^n = Y^n$ $(n \neq 0)$ implies $X = Y$ for $X, Y \in G$.*

An *Artinian group* is a group without infinite descending chains of subgroups. A group containing a subgroup of finite index which is a direct product of finitely many quasicyclic groups is called a *Chernikov group*. It is obvious that any Chernikov group is Artinian. Is the converse true (see [3])?

THEOREM 2 ([8],[10]). *There is an infinite group G in which every proper subgroup has a prime order and any two subgroups of the same order are conjugate. G can be effectively presented with two generators in such a way that there exist algorithms for recognizing both equality and conjugacy in G.*

Thus, of course, a positive answer to the problem of O. Yu. Schmidt is obtained (see [3]) about the existence of an infinite group in which all proper subgroups are finite and which is different from the quasicyclic groups. It disproves the conjecture [4] on the finiteness of a group with both increasing and decreasing chain conditions. It produces a negative answer to the question of A. G. Kurosh and S. N. Chernikov [4]: Does the minimum condition for subgroups imply the local finiteness of a group? The existence of uncountably many non-isomorphic groups like those in Theorems 1, 2 is proved in [11]. It also gives answers to some problems about the structure of a lattice of subgroups (see [16]) and to number of other questions raised in the group theory.

One of the features of author's papers is the proof of a large number of lemmas by simultaneous induction on a natural parameter i which, following the example of [7], we call the rank. The construction of the groups results from successive addition of defining relations of new ranks.

The group G, whose construction is the goal, is defined by a certain infinite sequence of defining relations, $G = \langle a, b \mid R_1 = 1, \ldots\rangle$, where the lengths $|R_i|$ increase very rapidly as $i \to \infty$.

For the definition of the word R_i we choose [8], [9] the, in a sense, minimal pair of words $\{A, B\}$ in the alphabet $a^{\pm 1}$, $b^{\pm 1}$ such that $AB \neq BA$ in $G_{i-1} = \langle a, b \mid R_1 = 1, \ldots, R_{i-1} = 1\rangle$, and the subgroup $\mathrm{gp}\{A, B\}$ is proper in G_{i-1}. In accordance with $A < B$ or $B < A$ the word R_i has the form (we simplify slightly the correct definition [8])

$$\begin{aligned} R_i &\,\underline{\circ}\, aBA^{k_{i,0}}BA^{k_{i,1}}B\ldots A^{k_{i,h_i}} \quad \text{or} \\ R_i &\,\underline{\circ}\, bBA^{k_{i,0}}BA^{k_{i,1}}B\ldots A^{k_{i,h_i}}, \end{aligned} \tag{2}$$

where $k_{i,j}$, h_i are some rapidly increasing parameters. In [8], [10] the definition of the word R_i is similar to (2) for even i. If i is an odd number then $R_i \,\underline{\circ}\, C_i^{p_i}$, where C_i is the minimal, in a sense, word whose order in G_{i-1} is infinite, and p_i is a certain large prime number. It is not very difficult to show that every proper subgroup of G is abelian (and that G is a periodical group in Theorem 2). But why is G a non-abelian or even a non-trivial group?

An important technical tool in proving this is using diagrams with long periodic words written on their boundaries. A *periodic word with the period C* is a subword of a certain power C^n of word C. We note that the study (without connections with combinatorial topology) of periodic words and relations was started by V. A. Tartakovskij in the forties. The powerful elaboration of new combinatorial ideas for the investigation of transformations of periodic words was obtained by P. S. Novikov and S. I. Adyan [6], [7] in solving the restricted Burnside problem for sufficiently large odd exponents n, $n \geqslant 665$ [1].

As a rule, we study periodic words with simple periods, called *simple* if they are cyclically irreducible and not proper powers in the free group. Two partitions $X \,\underline{\circ}\, X_1 \cdot X_2$ and $Y \,\underline{\circ}\, Y_1 \cdot Y_2$, where $|X|, |Y| \geqslant |A|$, of periodic words with a period A will be called *A-consistent* if X and Y are the subwords of a certain power of A such that $A^k \,\underline{\circ}\, C_1 \cdot XD_1 \,\underline{\circ}\, C_2 \cdot YD_2$ and the difference $|C_1X_1| - |C_2Y_1|$ is a multiple of $|A|$. The following properties of periodic words with simple periods are almost obvious:

(i) If $|X| \geqslant |A|$ and $A^k \,\underline{\circ}\, C_1XD_1 \,\underline{\circ}\, C_2XD_2$, then $C_1 \cdot XD_1$ and $C_2 \cdot XD_2$ are the A-consistent partitions.

(ii) If $|X| \geqslant |A| + |B|$ and X is both A- and B-periodic, then B is a cyclic shift of A.

(iii) If X is both A- and A^{-1}-periodic word, then $|X| < |A|$.

As a matter of fact, these properties of periodic words in the free group can be transferred with a modification to any group G_i. Among the new geometric ideas an important role is played by the idea of a band, that is, a 'narrow' and 'long' diagram with periodic boundaries:

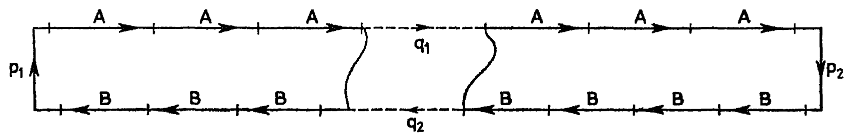


Fig. 2

Suppose $p_1q_1p_2q_2$ is the boundary path of the reduced diagram Δ of rank i (i.e. over the presentation G_i), $\varphi(q_1)$ is an A-periodic word, $\varphi(q_2)$ is a B-periodic word, and $|p_1|$, $|p_2|$, $|A|$, $|B| \ll |q_1|$, $|q_2|$, where A and B are simple in rank i, i.e., words which are minimal in their conjugacy classes of G_i and are not proper powers in G_i. Then:

(i) If $A \overset{\circ}{=} B^{-1}$, then the paths q_1 and q_2 have a common vertex o, which gives A-consistent partitions of $\varphi(q_1)$ and $\varphi(q_2)$.

(ii) B is conjugate to A^{-1} in G_i.

(iii) $A \neq B$ in G_i.

Our considerations of diagrams of rank i (a lot of details are avoided here) show that either a certain long subword of some defining word occurs in the boundary label of Δ or there exist two cells Π_1, Π_2 of the same rank whose boundary paths have many common edges. In the second case there exist bands 'between' Π_1 and Π_2. But the periodic structure of definig relations and the above properties of bands lead to a conclusion that the pair Π_1, Π_2 does not satisfy the condition in the definition of the reduced diagram. Thus, as in the Small Cancellation Theory, one deduces the existence, in any cyclically reduced consequence of $R_1, \ldots, R_i$, of a long piece of rank $j \leqslant i$. This enables us to prove, for example, that G_i is infinite, hence also G is infinite.

Other properties of G follow from a study of diagrams on a sphere or a torus. For instance, the nonexistence of reduced diagrams on a torus is used in the proof that any subgroup of G generated by two commuting elements is cyclic. It is known also (P. E. Schupp, see [5]) that conjugacy, as well as equality, has a natural geometric interpretation (with the help of annular diagrams). For the inductive proof of Theorems 1, 2 it is convenient to consider diagrams with an arbitrary number of 'holes'. The main additional difficulty in the proof of Theorem 2 is connected with the search, made geometrically, of elements of infinite order in the nonabelian subgroups of G_i.

Recently G. S. Deryabina has made some alterations in the proofs of Theorems 1, 2. An infinite nonabelian group with finite proper subgroups is called a Schmidt group. No 2-group is a Schmidt group (O. Yu. Schmidt [17]). However, for any prime $p \geqslant 3$, G. S. Deryabina has proved the existence of uncountably many non-isomorphic Schmidt p-groups with isomorphic lattices of subgroups (to appear). All the proper subgroups in these examples are cyclic and the maximal subgroups of the same order are conjugate.

There exist many equivalent definitions of the amenable group (see [2]). John von Neumann proved [19] that every locally solvable group is amenable and a group containing a free noncyclic subgroup is not amenable. The conjecture that any group without noncyclic free subgroups is amenable is attributed by Greenleaf [2] to John von Neumann [19]. We disprove it in [12].

THEOREM 3. *There exists a nonamenable group G such that every proper subgroup of G is cyclic.*

The proof is based on geometric estimations of the growth of the numbers d_n of words W with $|W| = n$ in the kernel of the presentation of Theorem 1 or 2.

As we have remarked above, the restricted Burnside problem for sufficiently large odd exponents has been solved in [7]. However, the considerable size and a complicated logical structure are the real obstacle to studying their paper. In [13] we succeeded in producing a new, relatively short, proof of the well-known theorem of Novikov–Adyan. It should be mentioned that our estimate for the exponent n is worse than that in [1].

THEOREM 4. *For any $m > 1$ and any $n > 10^{10}$ there exists an infinite m-generator group of exponent n (i.e. with the law $x^n \equiv 1$).*

A natural presentation of the given free Burnside group [13] can be described as follows. To define the ith relation we choose the shortest word C_i, in the alphabet $a_1^{\pm 1}, \ldots, a_m^{\pm 1}$, of infinite order in the group $\langle a_1, \ldots, a_m | \; R_1 = 1, \ldots, R_{i-1} = 1\rangle$. Then we put $R_i \mathrel{\overline{\circ}} C_i^n$.

Unlike [9], [10] the growth of $|R_i|$ cannot be arbitrarily rapid as $i \to \infty$. Moreover, the law $x^n \equiv 1$ causes the logarithmic growth of $|R_i|$. Therefore we have to find more delicate methods of dealing with bands. In place of this notion some other notions have appeared in [13]. Thus, we introduce the inductive notion of the contact diagram and the notion of a smooth section. An important Lemma 4.2 [13] states that the length of a smooth section of a path can be compared with another path between the same vertices.

The term 'Tarski-monster' appeared before the corresponding group was constructed in [14]. The problem lay in the existence of an infinite group in which every proper subgroup has the same prime order. The example [14] combines two strong finiteness conditions: it is the Schmidt group of restricted exponent.

THEOREM 5. *For every sufficiently large prime p ($p > 10^{75}$) there exists an infinite group G such that every proper subgroup of G has order p.*

New obstacles in [14] concern proving the cyclicity of every proper subgroup. Avoiding the details here, we note only that in [8]–[12] it is easy to achieve $|A^{k_{i,j}}| \gg |B|$ in relations (2) (even in the case $|A| \ll |B|$) by choosing $k_{i,j}$ sufficiently large. But it is obvious that in [14] all parameters should be bounded. The proof of Theorem 5 is based only on [13], although we use some ideas of [8], [9].

As the traditional Small Cancellations Theory, the author's method can be applied to relations additionally imposed on free products. For instance, starting from the group $A*B$, where $|B| = 3$, G. S. Deryabina (unpublished) has proved that every finite group A of odd order can be embedded in a Schmidt group with a number of additional properties.

In papers [8]–[14] we construct aspherical presentations (although in [13], [14] our notion of asphericity is weaker than in [5], the latter being connected with stronger restrictions for the reducibility of diagrams). Using the asphericity of presentations under consideration, we prove that F/N-module $N/[N, N]$, F being a free group and N the kernel of presentation (1), has no nontrivial relations (I. S. Ashmanov, A. Yu. Ol'šanskiĭ, unpublished). This fact enables us to construct abelian, in particular central, extensions of the above groups. For example, replacing the relations $R_1 = 1, R_2 = 1, \ldots$ of the group in Theorem 4 by $R_1 = R_2 = \ldots$ leads to the known group $A(m, n)$ due to S. I. Adyan, which answers a number of questions (see [1]). In particular, a factor-group of it is an answer to a problem of A. A. Markov on the existence of a countable nontopologized group [15]. A similar central extension of the group in Theorem 5 (now the centre is of order p) gives an answer to a known question of P. Hall (see, for instance, [21]): Is it true that the finiteness of a verbal subgroup V of a Noetherian group G implies finiteness of the index of the corresponding marginal subgroup V^*? The word that disproves the conjecture is x^p where p is the number in Theorem 5.

Thus, the new method has shown its power in constructing groups with a priori unknown but desired properties. There is no doubt that this method will also be used for solving some other group-theoretic problems. Finally, we mention an interesting paper of E. Rips [20], in which the theory of defining relations with special conditions has been elaborated. These conditions generalize the usual small cancellation hypotheses. For presentations with these conditions, E. Rips solves the word problem. The author of [20] has promised to apply, in the future, his theory to constructing groups solving some particular group-theoretic problems.

The author is grateful to Yu. A. Bakhturin for discussions during the writing.

References

[1] Адян С. И., *Проблема Бернсайда и тождества в группах*, Наука, Москва, 1975.

[2] Гринлиф Ф., *Инвариантные средние на топологических группах и их приложения*, Мир, Москва, 1973.

[3] Курош А. Г., *Теория групп*, ОГИЗ, Москва, 1944.

[4] Курош А. Г., Черников С. Н., Разрешимые и нильпотентные группы, *УМН* **2** (3) (1947), стр. 18–59.

[5] Линдон Р., Шупп П., *Комбинаторная теория групп*, Мир, Москва, 1980.

[6] Новиков П. С., О периодических группах, *ДАН СССР* **127** (1959), стр. 749–752.

[7] Новиков П. С., Адян С. И., Бесконечные периодические группы I, II, III, *Изв. АН СССР, сер. матем.* **32** (1968), 212–244, 251–254, 709–731.

[8] Ольшанский А. Ю., Бесконечные группы с циклическими подгруппами, *ДАН СССР* **245** (4) (1979), стр. 785–787.

[9] Ольшанский А. Ю., Бесконечная простая нетерова группа без кручения, *Изв. АН СССР, сер. матем.* **43** (6) (1979), стр. 1328–1393.

[10] Ольшанский А. Ю., Бесконечная группа с подгруппами простых порядков, *Изв. АН СССР, сер. матем.* **44** (2) (1980), стр. 309–321.

[11] Ольшанский А. Ю., О группах с циклическими подгруппами, *Доклады Болгарской АН* **32** (9) (1979), стр. 1165–1166.

[12] Ольшанский А. Ю., К вопросу о существовании инвариантного среднего на группе, *УМН* **35** (4) (1980), стр. 199–200.

[13] Ольшанский А. Ю., О теореме Новикова–Адяна, *Матем. сборник* **118** (2) (1982), стр. 203–235.

[14] Ольшанский А. Ю., Группы ограниченного периода с подгруппами простого порядка, *Алгебра и логика* **21** (5) (1982), стр. 553–618.

[15] Ольшанский А. Ю., Замечание о счетной нетопологизируемой группе, *Вестник МГУ, серия* 1, **3** (1980), стр. 103.

[16] Судзуки М., *Строение группы и строение структуры её подгрупп*, ИЛ, Москва, 1960.

[17] Шмидт О. Ю., *Избранные труды, Математика*, Москва, 1959.

[18] van Kampen E. R., On Some Lemmas in the Theory of Groups, *Amer. J. Math.* **55** (1933), pp. 268–273.
[19] von Neumann J., Zur allgemeinen Theorie des Maßes, *Fund. Math.* **13** (1929), pp. 73–116.
[20] Rips E., Generalized small cancellation theory and applications I. The word problem, *Israel J. Math.* **41** (1–2) (1982), pp. 1–146.
[21] Turner-Smith R. F., Marginal subgroup properties for outer commutator words, *Proc. London Math. Soc.* (*3*), **14** (1964), pp. 321–341.

MOSCOW STATE UNIVERSITY

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

CLAUS MICHAEL RINGEL

Indecomposable Representations of Finite-Dimensional Algebras

Let k be a field and A a finite-dimensional k-algebra (associative, with 1). We consider representations of A as rings of endomorphisms of finite-dimensional k-spaces, and thus A-modules, and we ask for a classification of such representations. More generally, we may consider the following problem: given an abelian category $\mathscr{C}$ and simple ($=$ irreducible) objects $E(1), \ldots, E(n)$ in $\mathscr{C}$, what are the objects in $\mathscr{C}$ of finite length with all composition factors of the form $E(1), \ldots, E(n)$. Problems of this kind arise naturally in many branches of mathematics, in particular, classification problems for linear representations of other algebraic structures (groups, Lie algebras, etc.) may be reinterpreted in this way. We will always assume that we know the simple A-modules $E(i)$, $1 \leqslant i \leqslant n$, and also their first extension groups $\mathrm{Ext}^1\big(E(i), E(j)\big)$, and thus the modules of length 2, and our aim is to study the indecomposable modules of greater length. Note that any (finite-dimensional) A-module can be written as a direct sum of indecomposable modules, and the Krull–Schmidt theorem asserts that these indecomposable direct summands, as well as their multiplicities, are uniquely determined. Besides the semisimple algebras (with all indecomposable modules being simple), there are other algebras with only finitely many (isomorphism classes of) indecomposable modules (they are said to be *representation finite*). However, there will usually be large families of (pairwise non-isomorphic) indecomposable modules, indexed over suitable algebraic varieties. As Drozd [16] has shown, for representation infinite algebras, there is a strict distinction between the tame and the wild representation type, the tame algebras being characterized by the property that there are at most one-parameter families of indecomposable modules. For a wild algebra, it seems difficult to obtain a complete classification of all indecomposable modules, since it would involve the (unsolved) problem of classifying pairs of square matrices with respect to joint similarity. A typical tame problem is the classification of all matrix pencils, which was solved by Kronecker in 1890. This result was needed for classifying pairs of symmetric bilinear forms, but it is also of interest for solving differential equations. Note that matrix pencils are nothing but modules over the four-dimensional algebra $\begin{bmatrix} k & k^2 \\ 0 & k \end{bmatrix}$.

For simplicity, we will usually assume that k is algebraically closed. Also, let A be connected (without central idempotents $\neq 0, 1$). Given a finite-dimensional algebra A, we denote by $\mathrm{K}_0(A)$ the Grothendieck group of all finite length A-modules modulo all exact sequences. The equivalence class in $\mathrm{K}_0(A)$ of an A-module M will be denoted by $\mathbf{dim}M$ and called its dimension vector. $\mathrm{K}_0(A)$ has a canonical basis given by the elements $\mathbf{dim}E(i)$, $1 \leqslant i \leqslant n$, and, in this way, we may identify $\mathrm{K}_0(A)$ with $\mathbf{Z}^n$. Evaluating $\mathbf{dim}M = \sum_{i=1}^{n} m_i \mathbf{dim}E(i)$, we see that the integer $m_i = (\mathbf{dim}M)_i$ is non-negative, it is just the Jordan–Hölder multiplicity of $E(i)$ in M (the multiplicity of $E(i)$ in a composition series of M; the invariance of m_i is usually called the Jordan–Hölder theorem). If all m_i are positive, M is called *sincere*, and A is called *sincere* provided there exists an indecomposable sincere A-module. The *Cartan matrix* C_A has as the (i, j)th entry the number $(\mathbf{dim}P(j))_i$, where $P(i)$ is a projective cover of $E(i)$. If C_A is invertible (for example, if gl.dim. $A < \infty$), consider the bilinear form $\langle x, y\rangle = xC_A^{-T}y^T$ on $\mathrm{K}_0(A)$. The quadratic form $\chi(x) = \langle x, x\rangle$ is called the *Euler characteristic* of A. Note that if X, Y are A-modules with proj.dim. $X < \infty$ or inj.dim. $Y < \infty$, then

$$\langle \mathbf{dim}X, \mathbf{dim}Y\rangle = \sum_{i \geqslant 0} (-1)^i \dim \mathrm{Ext}^i(X, Y).$$

It turns out that for some classes of algebras A the indecomposable A-modules are controlled by χ. The first result of this kind was Gabriel's theorem on hereditary representation finite algebras [17, 7]. In the same way, the hereditary tame algebras were first considered by Gelfand–Ponomarev, Donovan–Freislich and Nazarova [18, 15, 23, 14], and those which are wild by Kac [22]. We are going to outline that at least the representation finite and the tame cases can be well unterstood by considering the corresponding Auslander–Reiten quivers, and at the same time we want to consider some classes of tame algebras of global dimension 2.

1. The Auslander–Reiten quiver $\Gamma(A)$ of A

The present representation theory is based on some fundamental concepts due to Auslander and Reiten. These concepts are incorporated in the notion of the Auslander–Reiten quiver $\Gamma(A)$ of A. We are interested in the set of isomorphism classes $[M]$ of indecomposable A-modules M, and we consider it as the set of vertices of a translation quiver $\Gamma(A)$, with an arrow $[M]\to[N]$ provided there exists an irreducible map $M\to N$, and using the Auslander–Reiten translation τ.

1.1. (Auslander–Reiten [3]). For any indecomposable A-module Z, there exists a map $g\colon Y\to Z$ (unique up to an isomorphism) with the following three properties:

(i) g is not split epi,

(ii) if $g'\colon Y'\to Z$ is not split epi, then there is an $\eta\colon Y'\to Y$ satisfying $g' = \eta g$,

(iii) if $\zeta\colon Y\to Y$ satisfies $g = \zeta g$, then ζ is an automorphism.

Such a map $g\colon Y\to Z$ will be called a *sink map* for Z (Auslander–Reiten used the term "minimal right almost split map"). Dually, for any indecomposable A-module X, there exists (again unique up to an isomorphism) a *source map* $f\colon X\to Y'$.

A map $h\colon M\to N$ is said to be *irreducible* provided h is not a split map and has only trivial factorizations (i.e., $h = h'h''$ implies that h' is split mono, or h'' is split epi). Let M, N be indecomposable. Let $\operatorname{rad}(M, N)$ be the set of non-invertible maps $M\to N$, and

$$\operatorname{rad}^2(M, N) = \sum \operatorname{rad}(M, X)\operatorname{rad}(X, N)$$

where X runs through all indecomposable modules. Then $h\colon M\to N$ is irreducible if and only if $h \in \operatorname{rad}(M, N)\setminus\operatorname{rad}^2(M, N)$. We call $\operatorname{Irr}(M, N) = \operatorname{rad}(M, N)/\operatorname{rad}^2(M, N)$ the *bimodule* of irreducible maps. If $Y\to N$ is a sink map, or $M\to Y'$ a source map, then $\dim_k \operatorname{Irr}(M, N)$ measures the multiplicity of M in a direct decomposition of Y into indecomposable modules, and also the multiplicity of N in such a decomposition of Y'.

There is a strong interrelation between sink maps and source maps:

1.2. (Auslander–Reiten [3]). Let Z be an indecomposable A-module, and $g\colon Y\to Z$ a sink map. Then, either Z is projective, Y is its radical, and g the inclusion map, or else g is epi, $\tau Z := \operatorname{Ker}(g)$ is indecomposable, and the inclusion map $\tau Z\to Y$ is a source map.

We see that for Z indecomposable and not projective, we obtain (uniquely) an exact sequence

$$0\to\tau Z\xrightarrow{f} Y\xrightarrow{g} Z\to 0$$

with τZ also indecomposable, f a source map, g a sink map. These sequences are now called *Auslander–Reiten sequences* (Auslander–Reiten used the term "almost split sequences".) There is a direct recipe for constructing τZ for a given Z. Let $P_1 \xrightarrow{p} P_0 \to Z \to 0$ be a minimal projective presentation of Z, and let $\nu = D\mathrm{Hom}(-, {}_AA)$, then $\tau Z = \mathrm{Ker}(\nu p)$. In case the Cartan matrix C of A is invertible, the dimension vector of τZ is given as follows:

$$\mathbf{dim}\,\tau Z = (\mathbf{dim} Z)\Phi - (\mathbf{dim}\,\mathrm{Ker}\,p)\Phi + \mathbf{dim}\,\nu Z,$$

with $\Phi = -C^{-T}C$. Thus Φ measures the change of dimension vectors under the Auslander–Reiten translation [2].

Since $\Gamma(A)$ is locally finite, any component of $\Gamma(A)$ is either finite or countable, and finite only in case A is representation finite:

1.3. (Auslander [1]). *Assume that $\Gamma(A)$ has a component containing only modules of bounded length. Then A is representation finite and $\Gamma(A)$ is connected.*

This result of Auslander strengthens a theorem of Rojter which had established the first Brauer–Thrall conjecture. On the other hand, it also yields a method for showing that a given finite list of indecomposable A-modules is complete.

We will consider the possible shapes of components in the next sections. A component $\mathscr{C}$ of $\Gamma(A)$ will be said to have a *trivial modulation* provided $\dim_k \mathrm{Irr}(X, Y) \leqslant 1$ for all indecomposable modules X, Y in $\mathscr{C}$.

There are several ways of considering $\Gamma(A)$ as a two-dimensional simplicial complex, with 0-simplices the vertices of $\Gamma(A)$. Following Gabriel and Riedtmann, we take as 1-simplices both the pairs $([M], [N])$ with $[M] \to [N]$ in $\Gamma(A)$ and the pairs $([\tau Z], [Z])$ where Z is indecomposable and not projective, and as 2-simplices the triples $([\tau Z], [Y], [Z])$, again for Z indecomposable and not projective, and $[Y] \to [Z]$ in $\Gamma(A)$. In this way, we may speak of the underlying topological space of $\Gamma(A)$. As an example, the Auslander–Reiten quiver of the algebra of all upper triangular $n \times n$-matrices over k is just a large triangle, it will be denoted by $\Theta(n)$.

2. Directing modules and preprojective components

A *cycle* in the category A-mod of A-modules is a finite sequence $M_0, M_1, \ldots$ $\ldots, M_n = M_0$ of indecomposable modules satisfying $\mathrm{rad}(M_{i-1}, M_i) \neq 0$ for $1 \leqslant i \leqslant n$. Of course, any cyclic path in $\Gamma(A)$ gives a cycle in A-mod; however, there may exist additional cycles in A-mod, both containing modules composed of a single component and containing modules of different components. An indecomposable module M not contained in a cycle is called a *directing module*. An element x of $\mathrm{K}_0(A)$ will be called a *root* if $\chi(x) = 1$ and a *null root* if $\chi(x) = 0$.

2.1. If M is a directing module, then **dim** M is a positive root, and M is uniquely determined by $\mathbf{dim} M$.

2.2. The existence of a sincere directing A-module implies that gl. dim. $A \leqslant 2$.

2.3. A component $\mathscr{C}$ of $\Gamma(A)$ containing no cyclic path and such that any module in $\mathscr{C}$ is of the form $\tau^{-t}P$ for some indecomposable projective module P and some $t \geqslant 0$ is called a *preprojective component*. The modules belonging to a preprojective component are all directing modules, and they can be constructed inductively from the indecomposable summands of the radicals of the indecomposable projective modules as iterated cokernels. Given a preprojective component $\mathscr{C}$, its *orbit quiver* is a quiver with labelled arrows: its vertices are given by the τ-orbits in $\mathscr{C}$, or, equivalently, by the indecomposable projective modules in $\mathscr{C}$, and the number of arrows from $P(i)$ to $P(j)$ with label t is n_{ijt}, where $\mathrm{rad} P(i) = \bigoplus_{j,t} \left(\tau^{-t}P(j)\right)^{n_{ijt}}$. In case $\mathscr{C}$ is a preprojective component not containing indecomposable injective modules, then the full subcategory of all modules belonging to $\mathscr{C}$ is uniquely determined by k and the orbit quiver. (There is a general notion of preprojective modules due to Auslander and Smalø [4], the modules in a preprojective component being always preprojective.)

2.4. If A is a (finite-dimensional) hereditary algebra, then A is the path algebra of some finite quiver $\Delta(A)$. In this case, A has a unique preprojective component, and its orbit quiver is just $\Delta(A)$, with all labels being 0.

2.5. A representation finite (connected) algebra A is called *directed* provided there is no cycle in A-mod, or, equivalently, the indecomposable A-modules form a preprojective component which is finite. A quadratic form q on a free abelian group with a distinguished basis is said to be *integral* provided q takes only integer values, and q takes the value 1 on the elements of the distinguished basis; and q is said to be *weakly positive* provided q takes positive values on positive elements. If A is a directed algebra, then the Euler characteristic χ_A is obviously an integral quadratic form on $\mathrm{K}_0(A)$. In addition, if A is sincere, then χ_A is weakly positive, and **dim** furnishes a bijection between the indecomposable A-modules and the positive roots of χ_A [20]. (Actually, as Bongartz [9] has shown, in dealing with directed but not necessarily sincere algebras, one may replace the quadratic form χ by some truncated form χ' which is always weakly positive, integral and with **dim** defining a bijection between the indecomposable modules and the positive roots.) Now, an interesting theorem of Ovsienko [24] asserts that, for a positive root $\mathbf{x} = (x_1, \ldots, x_n)$ of a weakly positive integral quadratic form on $\mathbf{Z}^n$, all the coordinates satisfy $x_i \leqslant 6$, thus we see that the Jordan–Hölder multiplicities of an indecomposable module over a directed algebra are bounded from above by 6. (A different proof of this bound was given by Bongartz in [8].) It will be possible to classify the sincere directed algebras. In [8], Bongartz determined all such algebras with more than 336 simple modules: there are 24 different series of them, and its is interesting to note that for these algebras the Jordan–Hölder-multiplicities of the indecomposable modules are bounded even by 3. There are only finitely many additional sincere directed algebras, and the corresponding list should be furnished with the help of a computer. Finally, let us note that the possibilities for the orbit quivers of sincere directed algebras are rather restricted: the orbit quivers form a tree (Bautista–Larrion–Salmeron [6]) with at most four end-points (Bongartz [8]).

2.6. An algebra A is said to be *minimal representation infinite* provided A is not representation finite, but for any non-zero ideal I the algebra A/I is representation-finite. The minimal representation-infinite algebras having a preprojective component have been classified by Happel and Vossieck [21]. The underlying graph of the orbit quiver of the preprojective component is a Euclidean diagram ($\tilde{\boldsymbol{A}}_n$, $\tilde{\boldsymbol{D}}_n$, $\tilde{\boldsymbol{E}}_6$, $\tilde{\boldsymbol{E}}_7$, or $\tilde{\boldsymbol{E}}_8$). In case $\tilde{\boldsymbol{A}}_n$, the algebra is hereditary, in case $\tilde{\boldsymbol{D}}_n$, there are 4 different series of them, and the numbers of the isomorphism classes in the cases $\tilde{\boldsymbol{E}}_6$. $\tilde{\boldsymbol{E}}_7$, $\tilde{\boldsymbol{E}}_8$ are 56, 437, 3809, respectively.

The minimal representation infinite algebras with a preprojective component will be called *concealed algebras*.

3. Representation finite algebras

3.1. Given A, representation finite, one may construct the universal covering $\tilde{\Gamma}(A)$ of $\Gamma(A)$, say with the covering group $G(A)$ and the covering map π: $\tilde{\Gamma}(A) \to \Gamma(A)$. Bongartz and Gabriel [10] have shown that $\tilde{\Gamma}(A)$ itself is an Auslander–Reiten quiver, although usually not of a finite-dimensional algebra, but only of a locally bounded algebra $\tilde{A}$, called the *universal cover* of A. Given an indecomposable A-module M, we choose some indecomposable $\tilde{A}$-module $\tilde{M}$ with $\pi([\tilde{M}]) = [M]$ and consider the support algebra $\tilde{A}(\tilde{M}) = \tilde{A}/\langle e = e^2 \in \tilde{A} | e\tilde{M} = 0\rangle$. Then $\tilde{A}(\tilde{M})$ is a directed algebra with $\tilde{M}$ being a sincere $\tilde{A}(\tilde{M})$-module, and many properties of the A-module M can be dealt with considering $\tilde{M}$ as an $\tilde{A}(\tilde{M})$-module. Note that the simple $\tilde{A}(\tilde{M})$-modules are partitioned into subsets which correspond to the simple A-modules so that the Jordan–Hölder multiplicities of a simple A-module in M is given by the sum of the Jordan–Hölder multiplicities of the simple $\tilde{A}(\tilde{M})$-modules in the corresponding subset. Since Bretscher and Gabriel [12] have shown that $\tilde{A}$ and $G(A)$ can be constructed directly from A as soon as its quiver and a set of relations are known, one may use $\tilde{A}$ efficiently in order to determine $\Gamma(A)$. Namely, we construct $\Gamma(\tilde{A})$ inductively, using one-point extensions (note that this construction leads rather quickly to a periodic behaviour with respect to $G(A)$ since $(\mathbf{dim}\tilde{M})_i \leqslant 6$ for indecomposable $\tilde{M}$ and any i), and we obtain $\Gamma(A)$ from $\Gamma(\tilde{A})$ by factoring out the operation of $G(A)$ on $\Gamma(\tilde{A})$. We remark that Bongartz and Gabriel [10] have shown that $G(A)$ is always a free (non-abelian) group.

3.2. There is a general result concerning the Auslander–Reiten quiver of a representation finite algebra A. According to Bautista and Brenner [5], any element x of $\Gamma(A)$ has at most four direct predecessors and at most four direct successors, and the number four occurs only if one of the four modules is projective-injective. This may be rephrased as a statement concerning subquivers of $\Gamma(A)$ of form $\tilde{\boldsymbol{D}}_4$, and is a special case of a more general statement dealing with subquivers of $\Gamma(A)$ which are Euclidean diagrams: namely, one can determine a universal bound for their "replication length".

3.3. As a special case one obtains the following result of Riedtmann [25]: an indecomposable A-module M is called *periodic* provided $\tau^t M \approx M$ for some $t \geqslant 1$. The translation subquiver of $\Gamma(A)$ consisting of all periodic modules will be denoted by $\Gamma_p(A)$. Then, for a representation finite A, the components of $\Gamma(A)$ are of the form $\boldsymbol{Z}\Delta/G$, Δ being a Dynkin diagram ($\boldsymbol{A}_n$, $\boldsymbol{D}_n$, $\boldsymbol{E}_6$, $\boldsymbol{E}_7$, or $\boldsymbol{E}_8$) and G being a group of automorphism of $\boldsymbol{Z}\Delta$. For a representation infinite algebra A, it was shown in [19] that there is only one further possibility, namely $\Delta = \boldsymbol{A}_\infty$. We recall the construction of $\boldsymbol{Z}\Theta$ for an arbitrary quiver Θ: we start with disjoint copies of Θ indexed by $\boldsymbol{Z}$, say with elements (z, i) where $z \in \boldsymbol{Z}$, and i is a vertex of Θ. Then we add arrows $(x, i) \to (y, i+1)$ for any arrow $y \to x$ and any i, and define τ by $\tau(z, i) = (z, i-1)$. If the underlying graph $\bar{\Theta}$ of Θ is a tree, $\boldsymbol{Z}\Theta$ only depends on $\bar{\Theta}$.

4. Separating tubular families

A component of $\Gamma(A)$ with trivial modulation is called a *tube* provided it contains a cyclic path and its underlying topological space is of the form $S^1 \times \boldsymbol{R}_0^+$ (where S^1 is the 1-sphere, and $\boldsymbol{R}_0^+$ the set of non-negative real numbers). The *regular* tubes of rank r (where $r \geqslant 1$) are the components of the form $\boldsymbol{Z}\boldsymbol{A}_\infty/\langle\tau^r\rangle$. We want to outline a procedure for obtaining families of tubes $\mathcal{T}(\lambda)$ with $\lambda \in \boldsymbol{P}_1 k$. With any family of positive integers n_λ $(\lambda \in I)$, we associate its *type*: form the disjoint union of diagrams $\boldsymbol{A}_{n_\lambda}$ $(\lambda \in I)$, choose in any $\boldsymbol{A}_{n_\lambda}$ one end-point and identify these end-points in order to form a star. The type of the sequence $n_1, \ldots, n_r$ will be denoted by $\Delta_{n_1,\ldots,n_r}$. Given a family of tubes, its type will be the type of its rank function. A family of tubes $\mathcal{T}(\lambda)$, $\lambda \in I$, will be said to *separate* $\mathcal{P}$ from $\mathcal{Q}$, provided the remaining indecomposable A-modules fall into two classes $\mathcal{P}$, $\mathcal{Q}$ such that $\mathrm{Hom}(\mathcal{Q}, \mathcal{P}) = \mathrm{Hom}(\mathcal{Q}, \mathcal{T}(\lambda)) = \mathrm{Hom}(\mathcal{T}(\lambda), \mathcal{P}) = 0$ for all λ, and, moreover, such that, for any λ any map from a module in $\mathcal{P}$ to a module in $\mathcal{Q}$ factors through a direct sum of modules in $\mathcal{T}(\lambda)$.

We consider an algebra A which is a one-point extension of an algebra A_0, say $A = A_0[R] := \begin{bmatrix} A_0 & R \\ 0 & k \end{bmatrix}$ where R is an A_0-module. The A-modules can be written as triples (X_0, X_ω, γ), X_ω being a k-space and $\gamma\colon R \otimes X_\omega \to X_0$ being an A_0-linear map. The indecomposable A_0-module W_0 is said to be a *wing module* provided it belongs to a component with trivial modulation and such that any arrow $x \to [W_0]$ in $\Gamma(A_0)$ is contained in a full convex translation subquiver Θ_0 (a "wing") of $\Gamma(A_0)$ isomorphic to some $\Theta(n_x)$, $[W_0]$ being the projective-injective vertex of Θ_x. The type of W_0 is the type of the function n_x. A wing module W_0 is said to be *separating* provided the indecomposable A_0-modules not belonging to the wings fall into two classes $\mathcal{X}$, $\mathcal{Y}$ with $\mathrm{Hom}(\mathcal{Y}, \mathcal{X}) = \mathrm{Hom}(\mathcal{Y}, W_0) = \mathrm{Hom}(W_0, \mathcal{X}) = 0$, and such that any map from a module in $\mathcal{X}$ to a module in $\mathcal{Y}$ factors through a direct sum of copies of W_0. Finally, W_0 is said to be *dominated* by R, provided $\mathbf{dim} R = (\mathbf{dim} W_0)(T - \Phi^{-1})$ and, for any $0 \neq \varrho\colon R \to W_0$, the $A_0[R]$-module $W_0(\varrho) := (W_0, k, \varrho)$ satisfies proj. dim. $W_0(\varrho) \leqslant 1$.

4.1. Let W_0 be a sincere separating wing A_0-module of type Δ dominated by R. Let $A = A_0[R]$, let $w = \mathbf{dim} W_0 + \mathbf{dim}(0, k, o)$ in $\mathrm{K}_0(A)$, and let

$\iota_w = \langle w, -\rangle$. Let $\mathscr{P}_w$, $\mathscr{T}_w$, $\mathscr{Q}_w$ be the sets of indecomposable A-modules satisfying $\iota_w(\mathbf{dim}M) < 0$, $= 0$, > 0, respectively. Then $\mathscr{T}_w$ is a $\boldsymbol{P}_1 k$-family of regular tubes of type Δ separating $\mathscr{P}_w$ from $\mathscr{Q}_w$. Also, $\mathscr{T}_w$ is controlled by the restriction of χ_A to $\operatorname{Ker}\iota_w$ (for M in $\mathscr{T}_w$, $\mathbf{dim}M$ is a root or a null root; conversely, given a positive root a in $\operatorname{Ker}\iota_w$, there is a unique indecomposable M with $\mathbf{dim}M = a$, and given a positive null root a in $\operatorname{Ker}\iota_w$, there is a one-parameter family of such modules).

4.2. If Δ is neither a Dynkin diagram, nor a Euclidean diagram, then both $\mathscr{P}_w$ and $\mathscr{Q}_w$ are wild. Note that one may use (4.1) to construct tubular families of arbitrary type Δ.

4.3. Assume A_0 is a directed algebra and M a sincere maximal indecomposable module. Then M is dominated by some projective module R. Thus, if the τ-orbit of M is the only possible branching point in the orbit quiver, then we can apply (4.1) to $W_0 = M$.

In particular, if A_0 is hereditary and representation finite, then A_0 is directed and has a unique maximal module M, and we may apply (4.1) to $W_0 = M$. Let M be dominated by R. Then $A = A_0[R]$ is hereditary, too. The orbit type of A_0, and therefore the tubular type of A, is a Dynkin diagram Δ. In this case $\mathscr{P}_w$ is a single component, namely the preprojective component of A, and its orbit diagram is the corresponding extended diagram $\tilde{\Delta}$. Similarly, $\mathscr{Q}_w$ is the preinjective component of A, also with the orbit diagram $\tilde{\Delta}$. In this way, we obtain a full classification of the representations of a tame hereditary algebra [15, 23, 14]; actually, one may consider in the same way also the general case of an arbitrary base field as considered in the joint work of Dlab and the present author [14]. A similar structure theory holds for all concealed algebras. The indecomposable modules S belonging to the mouth of a tube (thus those with $\operatorname{Hom}(\tau S, S) = 0$) are said to be *simple regular*.

4.4. Let A_0 be a concealed algebra with a tubular family $\mathscr{T}_0(\lambda)$ of rank $r_0(\lambda)$, and S a simple regular A_0-module in some $\mathscr{T}_0(\varrho)$. Then $A = A_0[S]$ is called a *simple tubular extension* of A_0. The *extension type* of A is, by definition, the type of the function $r_A\colon \boldsymbol{P}_1 k \to \boldsymbol{N}$ with $r_A(\lambda) = r_0(\lambda)$, for $\lambda \neq \varrho$, and $r_A(\varrho) = r_0(\varrho)+1$. Inductively, one may define arbitrary tubular extensions, and their extension types (see [13, 27]). Note that all those algebras have global dimension $\leqslant 2$. The dual construction is that of a tubular coextension.

Let A be a tubular extension of a concealed algebra A_0, and let Δ

be the extension type of A. If Δ is neither Dynkin nor Euclidean, then A is wild. If Δ is Dynkin, then A has a preprojective component $\mathscr{P}$, namely the preprojective component of A_0. It has also a preinjective component $\mathscr{Q}$ containing all indecomposable injective modules with the orbit diagram $\tilde{\Delta}$, and the remaining indecomposable modules form a $\boldsymbol{P}_1 k$-family of (not necessarily regular) tubes separating $\mathscr{P}$ from $\mathscr{Q}$. Also, A-mod is controlled by χ. Thus, it remains to consider the case of a Euclidean Δ, so that $\Delta = \Delta_{n_1,\ldots,n_r}$, with $(n_1, \ldots, n_r) = (2, 2, 2, 2), (3, 3, 3), (4, 4, 2)$ or $(6, 3, 2)$.

4.5. Let A be a tubular extension of a concealed algebra A_0 of the Euclidean extension type Δ. Then A is also a tubular coextension of a concealed algebra A_∞. Let w_0 be the minimal positive null root of A_0, and w_∞ that of A_∞, and define $\iota_{\alpha,\beta} = \langle \alpha w_0 + \beta w_\infty, -\rangle$. We have the following components of $\Gamma(A)$:

(1) a preprojective component $\mathscr{P}_0$ (= the preprojective component of A_0),

(2) a separating tubular $\boldsymbol{P}_1 k$-family $\mathscr{T}_0$ (obtained from the tubular family of A_0 by ray insertions),

(3) for any $\gamma \in \boldsymbol{Q}^+$, a separating regular tubular $\boldsymbol{P}_1 k$-family $\mathscr{T}_\gamma$ of type Δ consisting of all indecomposable modules X with $\iota_{\alpha,\beta}(\mathbf{dim}\, X) = 0$, where $\gamma = \beta/\alpha$,

(2)* a separating tubular $\boldsymbol{P}_1 k$-family $\mathscr{T}_\infty$ (obtained from the tubular family of A_∞ by co-ray insertions),

(1)* a preinjective component $\mathscr{Q}_\infty$ (= the preinjective component of A_∞).

Also, the category A-mod is controlled by χ.

4.6. One may use this in order to obtain a corresponding result for subspace categories of vector space categories, in particular for the non-domestic partially ordered sets of finite growth. In fact, for these posets, the one-parameter families of indecomposable representations have been determined before by Nazarova and Zavadskij [30], and Zavadskij has independently constructed the remaining representations. The algebras considered here were studied by Brenner–Butler [11] and in [26]. The proof of the results in 4. are outlined in [27], they will appear in [28].

5. Addendum

We have tried to outline the use of the Auslander–Reiten quiver $\Gamma(A)$ for classifying indecomposable A-modules, and for getting some insight into the module category A-mod. Along these lines we have reported on some of the advances in representation theory in recent years. However, we should stress that this paper covers only a small portion of the present theory. For earlier results, we refer to the report of Rojter at the Helsinki Congress 1978 [29]. In the discussion of the material presented here, we have avoided the more technical notions (such as tilting functors) even if we are sure of their usefulness for a better understanding. Also, we have not discussed at all some of the major recent developments which would have justified a separate report. We only want to mention Riedtmann's classification of the representation finite selfinjective (= quasi-Frobenius) algebras, and the work of Rojter, and Bautista, Gabriel, Salmeron on the existence of a multiplicative basis for representation finite and minimal representation infinite algebras. In fact, it follows from the above work that for a basic representation finite algebra A, the Auslander–Reiten quiver $\Gamma(A)$ uniquely determines A, provided the characteristic of k is different from 2.

References

[1] Auslander M., Applications of Morphisms Determined by Objects, *Proc. Conf. Rep. Theory. M. Dekker* (1978), pp. 245–327.

[2] Auslander M. and Platzeck M. I., Representation Theory of Hereditary Artin Algebras, *Proc. Conf. Rep. Theory. M. Dekker* (1978), pp. 389–424.

[3] Auslander M. and Reiten I., Representation Theory of Artin Algebras III, *Comm. Algebra* **3** (1975), pp. 239–294, IV. *Comm. Algebra* **5** (1977), pp. 443–518.

[4] Auslander M. and Smalø S. O., Preprojective Modules over Artin Algebras, *J. Algebra* **66** (1980), pp. 61–122.

[5] Bautista R. and Brenner S., Replication Number for Non-Dynkin Sectional Subgraphs in Finite Auslander–Reiten Quivers and Some Properties of Weyl Roots, *Proc. London Math. Soc.* (3) **47** (1983), pp. 429–462.

[6] Bautista R., Larrion F. and Salmeron L., *On Simply Connected Algebras*, Preprint.

[7] Bernstein I. N., Gelfand I. M. and Ponomarev V. A., Coxeter Functors and Gabriel's Theorem, *Russian Math. Surveys* **28** (1973), pp. 17–32.

[8] Bongartz K., Treue einfach zusammenhängende Algebren I, *Comment. Math. Helv.* **57** (1982), pp. 282–330.

[9] Bongartz K., *Algebras and Quadratic Forms*, Preprint.

[10] Bongartz K. and Gabriel P., Coverings in Representation Theory, *Inv. Math.* **65** (1982), pp. 331–378.

[11] Brenner S. and Butler M. C. R., Generalizations of the Bernstein–Gelfand–Ponomarev Reflection Functors, *Proc. Ottawa Conf.*, Springer Lecture Notes in Mathematics, vol. 832 (1982), pp. 103–169.

[12] Bretscher O. and Gabriel P., The Standard Form of a Representation Finite Algebra, *Bull. Soc. math. France* **111** (1983), pp. 41–57.

[13] D'Este G. and Ringel C. M., Coherent Tubes, *Algebra* **87** (1984), pp. 150–201.

[14] Dlab V. and Ringel C. M., Indecomposable Representations of Graphs and Algebras, *Memoirs AMS* **173** (1976).
[15] Donovan P. and Freislich M. R., The Representation Theory of Finite Graphs and Associated Algebras, *Carleton Math. LN.* **5** (1973).
[16] Drozd J. A., Tame and Wild Matrix Problems, *Ottawa Conf.* Springer Lecture Notes in Mathematics 832 (1980), pp. 242–258.
[17] Gabriel P., Unzerlegbare Darstellungen I, *Manuscripta Math.* **6** (1972), pp. 71–103.
[18] Gelfand I. M. and Ponomarev V. A., Problems of Linear Algebra and Classification of Quadrupels in a Finite-Dimensional Vector Space, *Coll. Math. Soc. Bolyai* **5**, Tihany (1970), pp. 163–237.
[19] Happel D., Preiser U. and Ringel C. M., Vinberg's Characterization of Dynkin Diagrams Using Subadditive Functions with Application to DTr-Periodic Modules, *Ottawa Conf.*, Springer Lecture Notes in Mathematics 832 (1980), pp. 280–294.
[20] Happel D. and Ringel C. M., Tilted Algebras, *Trans. AMS* **274**, pp. 399–443.
[21] Happel D. and Vossieck D., Minimal Algebras of Infinite Representation Type with Preprojective Component, *Manuscripta Math.* **42** (1983), pp. 221–243.
[22] Kac V., Infinite Root Systems, Representations of Graphs, and Invariant Theory, *Invent. Math.* **56** (1980), pp. 57–92.
[23] Nazarova L. A., Representations of Quivers of Infinite Type, *Izv. Akad. Nauk SSSR, ser. mat.* **37** (1973), pp. 752–791.
[24] Ovsienko A. S., Integral Weakly Positive Forms. In: *Schurian Matrix Problems and Quadratic Forms*, Kiev (1978), pp. 3–17.
[25] Riedtmann Chr., Algebren, Darstellungsköcher, Überlagerungen und zurück, *Comment. Math. Helv.* **55** (1980), pp. 199–224.
[26] Ringel C. M., Tame Algebras, *Ottawa Conf.*, Springer Lecture Notes in Mathematics 831 (1980), pp. 137–287.
[27] Ringel C. M., *Separating Tubular Series*, Sem. Malliavin, Springer Lecture Notes in Mathematics 1029 (1983), pp. 134–158.
[28] Ringel C. M., *Tame Algebras and Root Systems*, to appear.
[29] Rojter A. V., Matrix Problems, *Proceedings Intern. Congress Math.*, Helsinki 1978.
[30] Zavadskij A. G. and Nazarova L. A., Partially Ordered Sets of Tame Type. In: *Matrix Problems*, Kiev (1977), pp. 122–143.

FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
D-4800 BIELEFELD 1
FRG

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

CHRISTOPHE SOULÉ

K-théorie et zéros aux points entiers de fonctions zêta

Borel a montré que le rang des groupes de K-théorie de l'anneau des entiers d'un corps de nombres est donné par l'ordre des zéros aux entiers négatifs de sa fonction zêta. Par ailleurs, Birch et Swinnerton-Dyer ont conjecturé que le rang du groupe des points d'une courbe elliptique E sur $\boldsymbol{Q}$ est l'ordre du zéro de sa fonction L en son centre de symétrie. Comme $K_0(E) = \boldsymbol{Z} \oplus E(\boldsymbol{Q}) \oplus \boldsymbol{Z}$, Bloch s'est demandé s'il existe un énoncé (conjectural) qui englobe ces deux situations [2]. Ceci a été réalisé par Beilinson [1]. On propose ici une version légèrement différente de celle de Beilinson, qui relierait le zéro au point entier j de la fonction zêta d'un schéma quasi-projectif sur $\boldsymbol{Z}$ à la caractéristique d'Euler–Poincaré de la partie de poids j (pour les opérations d'Adams) de la K-théorie des faisceaux cohérents sur ce schéma.

Je remercie B. Gross, S. Lichtenbaum et J.-L. Loday pour leur aide.

1. Les "poids" de la K'-théorie d'un schéma

1.1. Soient S un schéma régulier noethérien irréductible de dimension de Krull finie, et $\mathscr{V}$ la catégorie des schémas quasi-projectifs sur S. Si X est un objet de $\mathscr{V}$, on note $\dim(X)$ sa dimension de Krull et $M(X)$ la catégorie exacte des $\mathcal{O}_X$-modules cohérents. Quillen [19] associe à $M(X)$ une catégorie $QM(X)$ dont les objets sont ceux de $M(X)$ et dont les flèches sont définies à l'aide des suites exactes de $M(X)$. Si $BQM(X)$ est l'ensemble simplicial classifiant de cette catégorie, il pose

$$K'_m(X) = \pi_{m+1} BQM(X), \qquad m \geqslant 0.$$

Si X_{red} est le schéma réduit associé à X, on a $K'_m(X) = K'_m(X_{\mathrm{red}})$. Quand

X est régulier, on a $K'_m(X) = K_m(X)$, où $K_m(X)$ est défini comme $K'_m(X)$ en remplaçant $M(X)$ par la catégorie $P(X)$ des $\mathcal{O}_X$-modules cohérents localement libres.

Supposons que $X \to M$ est une immersion fermée de X dans un schéma de $\mathcal{V}$ lisse sur S. On pose

$$K_m^X(M) = \pi_{m+1}\big(\text{fibre}\big(BQP(M) \to BQP(M-X)\big)\big).$$

On a, d'après [19], un isomorphisme canonique $K'_m(X) \simeq K_m^X(M)$. Brown et Gersten [7] ont montré par ailleurs que $K_m^X(M)$ est le groupe de cohomologie à support:

$$K_m^X(M) = H_X^{-m}(M, \boldsymbol{Z} \times \boldsymbol{BGL}^+),$$

où $\boldsymbol{Z} \times \boldsymbol{BGL}^+$ est le faisceau simplicial associé au préfaisceau

$$U \mapsto \boldsymbol{Z} \times BGL\big(\Gamma(U, \mathcal{O}_M)^+\big)$$

(pour la définition de $BGL(A)^+$ pour un anneau unitaire A, voir [18] et [16]).

1.2. Le produit tensoriel de modules induit un produit

$$K_m^X(M) \times K_n^X(M) \to K_{m+n}^X(M).$$

Les opérations de puissances extérieures fournissent des applications

$$\lambda^k\colon K_m^X(M) \to K_m^X(M).$$

Celles-ci s'obtiennent en associant aux puissances extérieures des endomorphismes du faisceau $\boldsymbol{Z} \times \boldsymbol{BGL}^+$. On définit les opérations d'Adams ψ^k, $k \geqslant 1$, par la relation de récurrence

$$\psi^k(x) = \lambda^1(x)\psi^{k-1}(x) - \lambda^2(x)\psi^{k-2}(x) + \ldots + (-1)^{k-2}\lambda^{k-1}(x)\psi^1(x) + \\ + (-1)^{k-1}k\lambda^k(x),$$

si $x \in K_0^X(M)$, et par $\psi^k(x) = (-1)^{k-1}k\lambda^k(x)$, si $x \in K_m^X(M)$ et $m > 0$. Ce sont des endomorphismes de l'anneau gradué $\bigoplus_m K_m^X(M)$, qui vérifient la relation $\psi^k \circ \psi^{k'} = \psi^{kk'}$. Si i est un entier positif, on désigne par $K_m^X(M)^{(i)}$ le sous-espace de $K_m^X(M) \otimes \boldsymbol{Q}$ où toute opération ψ^k agit par multiplication par k^i. On montre qu'on a alors une décomposition en somme directe

$$K_m^X(M) \otimes \boldsymbol{Q} = \bigoplus_{i=\dim(M)-\dim(X)}^{m+\dim(M)} K_m^X(M)^{(i)}.$$

1.3. Supposons que le schéma M est irréductible. Si $j \in \boldsymbol{Z}$, on désigne par $K'_m(X)_{(j)}$ le sous-espace de $K'_m(X) \otimes \boldsymbol{Q}$ correspondant à $K^X_m(M)^{(\dim(M)-j)}$ par l'isomorphisme $K'_m(X) \otimes \boldsymbol{Q} \simeq K^X_m(M) \otimes \boldsymbol{Q}$. Un théorème de Riemann–Roch, analogue à ceux de [14], [12], et [20], montre que $K'_m(X)_{(j)}$ ne dépend pas du choix du plongement $X \to M$ comme ci-dessus, et que ce groupe est covariant (resp. contravariant) pour les morphismes projectifs (resp. étales) de $\mathscr{V}$.

Si $Y \to X$ est une immersion fermée de $\mathscr{V}$ et $U = X - Y$, on a une suite exacte longue

$$\ldots \to K'_m(Y)_{(j)} \to K'_m(X)_{(j)} \to K'_m(U)_{(j)} \to K'_{m-1}(Y)_{(j)} \to \ldots$$

1.4. Quand S est le spectre d'un corps, on montre que $K'_0(X)_{(j)}$ coïncide avec $CH_j(X) \otimes \boldsymbol{Q}$, où $CH_j(X)$ désigne le groupe d'homologie de Chow [11] des cycles de dimension j modulo l'équivalence linéaire. Dans [15] Landsburg introduit des groupes de Chow supérieurs $CH^i(X, m)$ (égaux à $CH^i(X)$, si $m = 0$). Quand X est affine et lisse sur un corps de caractéristique strictement supérieure à deux, on peut montrer que $K_m(X)^{(i)} = CH^i(X, m) \otimes \boldsymbol{Q}$.

2. Fonction zêta

2.1. On suppose désormais que $S = \operatorname{Spec}(\boldsymbol{Z})$. Soient X un schéma de $\mathscr{V}$ et $d = \dim(X)$ sa dimension de Krull. Notons $|X|$ l'ensemble des points fermés de X, et $N(x)$ l'ordre du corps résiduel de $x \in |X|$. La fonction zêta de X est définie par le produit infini

$$\zeta_X(s) = \prod_{x \in |X|} \left(1 - N(x)^{-s}\right)^{-1}.$$

Ce produit converge pour tout nombre complexe s tel que $\operatorname{Re}(s) > d$. C'est une conjecture "standard" que $\zeta_X(s)$ admet un prolongement analytique à tout le plan complexe.

2.2. Conjecture. *Soient X un schéma quasi-projectif sur $\boldsymbol{Z}$ et $j \in \boldsymbol{Z}$ un entier. Alors*

(i) *le groupe $K'_m(X)_{(j)}$ est nul pour presque tout entier m,*
(ii) *la dimension de $K'_m(X)_{(j)}$ sur $\boldsymbol{Q}$ est finie,*
(iii) *l'ordre du zéro de $\zeta_X(s)$ au point $s = j$ est égal à*

$$\sum_{m \geqslant 0} (-1)^{m+1} \dim_{\boldsymbol{Q}} K'_m(X)_{(j)}.$$

La conjecture (i) a été faite indépendamment par Beilinson, (ii) est dû à Bass. Dans (iii) un zéro d'ordre négatif est un pôle.

2.3. Si $j > d$, la conjecture est vraie (les deux termes de l'égalité sont nuls). Si $j = d$, elle est également vérifiée: la fonction $\zeta_X(s)$ admet un prolongement analytique au demi-plan $\mathrm{Re}(s) > d-(1/2)$ et possède en $s = d$ un pôle d'ordre le nombre de composantes de X de dimension d [21], i.e. le rang de $K'_0(X)_{(d)} = CH_d(X)\otimes \mathbf{Q}$. Par ailleurs, $K'_m(X)_{(d)} = 0$ si $m > 0$.

Si $j = d-1$ et si X est régulier et irréductible, on a $K'_0(X)_{(d-1)} = H^1(X, \mathbf{G}_m)\otimes \mathbf{Q}$, $K'_1(X)_{(d-1)} = H^0(X, \mathbf{G}_m)\otimes \mathbf{Q}$, et $K'_m(X)_{(d-1)} = 0$ si $m \geqslant 2$. La conjecture 2.2 dans ce cas a été formulée par Tate ([26], BSD+II). Elle implique celle de Birch et Swinnerton-Dyer.

2.4. On a $\zeta_X(s) = \zeta_{X_{\mathrm{red}}}(s)$, donc la conjecture 2.2 ne dépend que du schéma réduit associé à X. Par ailleurs, si $Y \to X$ est une immersion fermée de $\mathscr{V}$ et $U = X - Y$, on a $\zeta_X(s) = \zeta_Y(s)\zeta_U(s)$. Il en résulte (grâce à 1.3) que si la conjecture est vraie pour deux des schémas X, Y et U, elle est vraie pour le troisième. En particulier, si la conjecture est vraie pour les schémas réguliers, elle est vraie pour tous les schémas de $\mathscr{V}$.

3. Variétés sur un corps fini

3.1. Soit X un produit de courbes et de variétés abéliennes sur un corps fini. Si $\dim(X) \leqslant 3$, la conjecture est vraie [23].

3.2. Quand X est lisse, projective et irréductible sur un corps fini, Tate a conjecturé que l'ordre du pôle de $\zeta_X(s)$ en $s = d-i$ est la dimension des classes de cycles algébriques de codimension i en cohomologie l-adique [26]. Dans les quelques cas décrits en 3.1 on montre, outre cela, que $K_m(X)\otimes \mathbf{Q} = 0$ si $m > 0$ et que l'équivalence cohomologique implique l'équivalence linéaire (modulo torsion).

4. Variétés sur $\mathbf{Q}$, approchetr anscendante

4.1. Soit V une variété projective lisse de dimension $d-1$ sur $\mathbf{Q}$. A cette variété est associée une fonction $\zeta_V(s)$ qui est un produit alterné

$$\zeta_V(s) = \prod_{k=1}^{k=2d-1} L_{k-1}(V, s)^{(-1)^{k-1}}.$$

Chaque fonction $L_{k-1}(V, s)$ est définie par un produit eulérien à partir de l'action de $\mathrm{Gal}(\overline{\boldsymbol{Q}}/\boldsymbol{Q})$ sur la cohomologie de $V\otimes\overline{\boldsymbol{Q}}$ en degré $k-1$ [22]. On pense (loc.cit.) que ce produit converge et est non nul quand $\mathrm{Re}(s) > (k+1)/2$, et que $L_{k-1}(V, s)$ admet un prolongement analytique au plan complexe et une équation fonctionnelle reliant ses valeurs en s et $k-s$.

Supposons que V admet un modèle régulier complet X sur $\boldsymbol{Z}$: $V = X\otimes_{\boldsymbol{Z}}\boldsymbol{Q}$. Alors l'image de $K_m(X)^{(i)}$ dans $K_m(V)^{(i)}$ est indépendante du choix de X [1]. On la note $K_m(V_{\boldsymbol{Z}})^{(i)}$. Beilinson conjecture que si $k<2i$, la dimension sur $\boldsymbol{Q}$ de $K_m(V_{\boldsymbol{Z}})^{(i)}$ est égale à l'ordre du zéro de $L_{k-1}(V, s)$ au point $s = k-i$.

Ceci est compatible à 2.2, au moins dans le sens suivant. Supposons que V est irréductible. Les facteurs eulériens des fonctions $\zeta_V(s)$ et $\zeta_X(s)$ sont égaux à $\zeta_{X\otimes\boldsymbol{F}_p}(s)$ quand X est lisse en p. Par suite, l'ordre du zéro de $\zeta_V(s)$ et de $\zeta_X(s)$ est le même en presque tout point entier s. Il en est de même de $L_{k-1}(V, s)$ et $L_{2d-k-1}(V, s+d-k)$ (dualité de Poincaré). Par ailleurs, on tire de 1.3 la suite exacte

$$\ldots\to K'_m(X)_{(j)}\to K_m(V_{\boldsymbol{Z}})^{(d-j)}\to \bigoplus_p K'_{m-1}(X\otimes\boldsymbol{F}_p)_{(j)}\to K'_{m-1}(X)_{(j)}\to\ldots$$

D'après 3.2, on peut penser que, pour presque tout entier j, les groupes $K'_m(X\otimes\boldsymbol{F}_p)_{(j)}$ sont nuls, donc $K'_m(X)_{(j)} = K_m(V_{\boldsymbol{Z}})^{(d-j)}$ (c'est vrai si $d\leqslant 2$). Donc (du moins si $d\leqslant 2$), 2.2 et la conjecture de Beilinson ci-dessus sont équivalentes pour presque toutes les valeurs entières de s.

4.2. Soit $V(\boldsymbol{C})$ l'ensemble des points complexes de V. Notons $H^{k-1}\big(V\otimes\boldsymbol{R}, \boldsymbol{R}(i-1)\big)$ la partie du groupe de cohomologie singulière $H^{k-1}\big(V(\boldsymbol{C}), \boldsymbol{R}\big)$ où la conjugaison complexe opère (via son action sur $V(\boldsymbol{C})$) par multiplication par $(-1)^{i-1}$. Soit φ_i l'application composée:

$$H^{k-1}_{DR}(V\otimes\boldsymbol{R})\to H^{k-1}_{DR}(V\otimes C) = H^{k-1}\big(V(\boldsymbol{C}), \boldsymbol{C}\big)\xrightarrow[\alpha_i]{} H^{k-1}\big(V(\boldsymbol{C}), \boldsymbol{R}\big),$$

où α_i est donnée sur les coefficients par la partie réelle (resp. imaginaire) multipliée par π^{i-1}, selon que i est impair (resp. pair). L'image de φ_i est $H^{k-1}\big(V\otimes\boldsymbol{R}, \boldsymbol{R}(i-1)\big)$. On pose:

$$R_{i,k} = H^{k-1}\big(V\otimes\boldsymbol{R}, \boldsymbol{R}(i-1)\big)/\varphi_i\big(F^iH^{k-1}_{DR}(V\otimes\boldsymbol{R})\big),$$

où F^i désigne la filtration de Hodge.

Bloch dans le cas des courbes [2] et Beilinson dans le cas général ont défini, pour $k<2i-1$, des morphismes régulateurs

$$\varrho_{i,k}\colon K_{2i-k}(V)^{(i)}\to R_{i,k}.$$

La dimension $n_{i,k}$ de l'espace vectoriel réel $R_{i,k}$ est égale à l'ordre du zéro de $L_{k-1}(V, s)$ en $s = i-k$, tel qu'on peut le prévoir par les conjectures sur l'équation fonctionnelle [22]. Beilinson pense que $\varrho_{i,k}$ induit un isomorphisme de $K_{2i-k}(V_{\boldsymbol{Z}})^{(i)}$ avec une $\boldsymbol{Q}$-forme de $R_{i,k}$.

Par ailleurs, $H^{k-1}(V \otimes \boldsymbol{R}, \boldsymbol{R}(i-1))$ et $F^i H^{k-1}_{DR}(V \otimes \boldsymbol{R})$ ont une $\boldsymbol{Q}$-structure naturelle, et on peut donc calculer le volume de $\varrho_{i,k}(K_{2i-k}(V_{\boldsymbol{Z}})^{(i)})$ par rapport à cette structure: c'est la classe d'un nombre réel modulo la multiplication par un élément de $\boldsymbol{Q}^*$. Si $c(i, k)$ est le premier coefficient non nul du développement de Taylor de $L_{k-1}(V, s)$ en $s = k-i$, Beilinson conjecture que le volume ci-dessus est la classe de $c(i, k)$ (modulo $\boldsymbol{Q}^*$) si $k < 2i-1$. Ceci peut aussi s'exprimer comme un résultat sur le nombre $L_{k-1}(V, i)$ (on notera que i est dans le demi-plan de convergence).

4.3. Les conjectures 2.2, 4.1, et 4.2 sont prouvées par Borel [5], [6], quand V est le spectre d'un corps de nombres. Il en résulte que 4.1 est vrai pour une surface rationnelle ou une variété de Severi–Brauer.

4.4. Quand V est une courbe elliptique à multiplication complexe (resp. une courbe modulaire), Bloch [2] (resp. Beilinson [1]) exhibent dans $K_2(V_{\boldsymbol{Z}})^{(2)}$ un sous-espace Ω tel que $\dim(\Omega) \geqslant n_{2,2}$. De plus, le volume de $\varrho_{2,2}(\Omega)$ est la classe de $c(2,2)$.

Des calculs numériques de Bloch et Grayson pour des courbes elliptiques sans multiplication complexe ont mis en évidence qu'il est crucial de considérer $K_{2i-k}(V_{\boldsymbol{Z}})^{(i)}$ et non pas $K_{2i-k}(V)^{(i)}$. D'autres exemples ont été étudiés par les auteurs cités et par D. Ramakrishnan.

4.5. Beilinson [1] et Bloch [3] pensent que l'ordre du zéro de $L_{2i-1}(V, s)$ en $s = i$ est le rang du groupe des cycles de codimension i sur V qui sont homologiquement triviaux, modulo l'équivalence linéaire. La valeur de $c(i, 2i)$ ferait intervenir une généralisation de la hauteur donnée dans [1] et [4]. Pour $i = 1$ on retrouve l'énoncé de Birch et Swinnerton-Dyer. Rappelons que Tate ([26], § 4) pense que $L_{2i}(s)$ a en $s = i+1$ un pôle d'ordre égal au rang des cycles de codimension i sur V, modulo l'équivalence homologique; pour la valeur de $c(i, 2i-1)$ et $c(i, 2i+1)$, voir [1].

Enfin la théorie s'étend, comme dans [9], des variétés sur $\boldsymbol{Q}$ aux motifs à coefficients dans un corps de nombres [1].

5. Variétés sur $\boldsymbol{Q}$, approche p-adique

5.1. Soient p un nombre premier et X un schéma quasi-projectif régulier sur $\boldsymbol{Z}$ (ce qui suit s'étend, comme au paragraphe 1, au cas singulier). Un approche plus indirecte du lien entre la K-théorie de X et sa fonction zêta consiste à relier ces deux notions à la cohomologie p-adique de $X' = X \otimes \mathbf{Z}[1/p]$.

Le lien entre fonction zêta et cohomologie p-adique est bien connu quand X est défini sur un corps fini (de caractéristique différente de p; Grothendieck, Deligne). En caractéristique zéro il est donné par la théorie d'Iwasawa [8], et n'est totalement compris que dans le cas où V est le spectre d'un corps de nombres totalement réel abélien sur $\mathbf{Q}$ (Mazur–Wiles).

Un lien entre K-théorie et cohomologie p-adique est donné par des classes de Chern

$$c_{i,k,n}\colon K_m(X, \mathbf{Z}/p^n)^{(i)} \to H^k_{\text{ét}}\big(X', \mathbf{Z}/p^n(i)\big), \qquad m+k = 2i,\ i \geqslant 1,$$

où $K_m(X, \mathbf{Z}/p^n)^{(i)}$ désigne la partie des groupes d'homotopie à coefficients $\mathbf{Z}/p^n$ de $BQP(X)$ où les opérations d'Adams ψ^k opèrent par multiplication par k^i. Notons $K_m(X, \mathbf{Q}_p/\mathbf{Z}_p)^{(i)}$ (resp. $H^k\big(X', \mathbf{Q}_p/\mathbf{Z}_p(i)\big)$) la limite inductive sur n des groupes $K_m(X, \mathbf{Z}/p^n)^{(i)}$ (resp. $H^k\big(X', \mathbf{Z}/p^n(i)\big)$) et $c_{i,k}$ la limite inductive des morphismes $c_{i,k,n}$.

Si A est un groupe de p-torsion, appelons $\dim_p A$ la dimension sur le corps à p éléments du noyau de la multiplication par p dans le sous-groupe divisible maximal de A. Si $K_m(X)$ et $K_{m-1}(X)$ sont des groupes de type fini, on a $\dim_p K_m(X, \mathbf{Q}_p/\mathbf{Z}_p)^{(i)} = \dim_{\mathbf{Q}} K_m(X)^{(i)}$.

5.2. On peut montrer que, pour p fixé, on a

$$\dim_p H^k\big(X', \mathbf{Q}_p/\mathbf{Z}_p(i)\big) = n_{i,k}$$

pour presque tout entier i [24]. En combinant les résultats de [10] et [27], on peut montrer que, si i est assez grand, le conoyau de

$$K_m(X', \mathbf{Q}_p/\mathbf{Z}_p)^{(i)} \to H^k\big(X', \mathbf{Q}_p/\mathbf{Z}_p(i)\big), \qquad m+k = 2i,$$

est d'exposant fini. La conjecture de Lichtenbaum–Quillen [18] impliquerait que, si i est assez grand, ces deux groupes ont même $\dim_p$.

5.3. Quand V est le spectre d'un corps de nombres et $i \geqslant 2$, le noyau et le conoyau de $c_{i,k}$ sont des groupes finis [25]. Il en résulte, grâce au théorème de Borel (4.3, [5]), que

$$\dim_p H^k\big(X', \mathbf{Q}_p/\mathbf{Z}_p(i)\big) = n_{i,k}, \qquad i \geqslant 2,$$

un résultat que, à ma connaissance, on ne sait montrer sans K-théorie que dans le cas d'une extension abélienne de $\mathbf{Q}$.

5.4. Merkurjev et Suslin ont montré que le symbole galoisien

$$c_{2,2,n}\colon K_2(F, \boldsymbol{Z}/p^n)^{(2)} \to H^2\big(F, \boldsymbol{Z}/p^n(2)\big)$$

est un isomorphisme pour tout corps F (contenant $1/p$) [17]. Si X est le modèle de Néron régulier complet d'une courbe elliptique V à multiplication complexe sur $\boldsymbol{Q}$, on en déduit (en prenant $F = \boldsymbol{Q}(V)$) que

$$\dim_p K_2(X, \boldsymbol{Q}_p/\boldsymbol{Z}_p)^{(2)} = \dim_p H^2\big(X', \boldsymbol{Q}_p/\boldsymbol{Z}_p(2)\big).$$

Quand $V = A(-p)$ est une des $\boldsymbol{Q}$-courbes de Gross [13] ou quand p est régulier au sens de Yager [28], on montre que

$$\dim_p H^2\big(X', \boldsymbol{Q}_p/\boldsymbol{Z}_p(2)\big) = 1 \ \ (= n_{2,2}).$$

Bibliographie

[1] Beilinson A. A., Higher Regulators and Values of L-Functions of Curves, *Funct. Anal. Appl.* **14** (2) (1980), pp. 116–117, et *Régulateurs supérieurs et valeurs de fonctions L* (1982), 90 pp.

[2] Bloch S., Algebraic K-Theory and Zeta Functions of Elliptic Curves, *Proc. Int. Congress of Mathematicians, Helsinki* (1978), pp. 511–515, et *Lectures on Algebraic Cycles,* Duke Univ. Math. Series 4 (1980).

[3] Bloch S., *Algebraic Cycles and Values of L-functions,* prépublication (1982), 32 pp.

[4] Bloch S., *Height Pairings for Algebraic Cycles,* prépublication (1983), 44 pp.

[5] Borel A., Stable Real Cohomology of Arithmetic Groups, *Ann. Sci. École Norm. Sup. Sér. 4* **7** (1974), pp. 235–272.

[6] Borel A., Cohomologie de SL_n et valeurs de fonctions zêta, *Ann. Scuola Norm. Sup. Pisa* **4** (1977), pp. 613–636.

[7] Brown K. and Gersten S. M., Algebraic K-Theory as Generalized Sheaf Cohomology, *Lect. Notes in Math.* 341, Springer-Verlag (1973), pp. 266–292.

[8] Coates J. and Lichtenbaum S., On l-adic Zeta Functions, *Ann. of Math.* **98** (1973), pp. 498–550.

[9] Deligne P., Valeurs de fonctions L et périodes d'intégrales, *Proc. Symp. Pure Math.* **33**, Vol. 2 (1979), pp. 313–346.

[10] Dwyer W. and Friedlander E., Étale K-Theory and Arithmetic, *Bull. Amer. Math. Soc.* (6) **3** (1982), pp. 453–455. Article détaillé en préparation.

[11] Fulton W., Rational Equivalence on Singular Varieties, *Publ. Math. IHES* **45**, (1975), pp. 147–167.

[12] Gillet H., Riemann–Roch Theorems for Higher Algebraic K-Theory, *Advances in Math.* **40** (3) (1981), pp. 203–289.

[13] Gross B., Arithmetic on Elliptic Curves with Complex Multiplication, *Lect. Notes in Math.* 776, Springer-Verlag (1980).

[14] Grothendieck A. *et al.*, Théorie des intersections et théorème de Riemann–Roch, *SGA* 6, *Lect. Notes in Math.* 225, Springer-Verlag (1971).

[15] Landsburg S. E., *Relative Cycles and Algebraic K-Theory,* prépublication (1983).

[16] Loday J.-L., K-théorie algébrique et représentations de groupes, *Ann. Sci. École Norm. Sup. Sér. 4* **9** (1976), pp. 309–377.

[17] Merkurjev A. S. et Suslin A. A., K-cohomologie des variétés de Severi–Brauer et symbole des restes normiques (en russe), *Izv. Akad. Nauk SSSR Ser. Mat.* **46** (1982), pp. 1011–1046.

[18] Quillen D., Higher Algebraic K-Theory, *Proc. Int. Congress of Mathematicians, Vancouver* (1974), pp. 171–176.

[19] Quillen D., Higher Algebraic K-Theory I, *Lect. Notes in Math.* 341, Springer-Verlag (1973), pp. 85–147.

[20] Shechtman V., Riemann–Roch for Higher K_i and a Partial Degeneration of the Atiyah–Hirzebruch Spectral Sequence (en russe), *Uspekhi Mat. Nauk* **35** (6) (1980), pp. 179–180.

[21] Serre J.-P., Zeta and L-Functions, dans: *Arithmetical Algebraic Geometry*, Harper and Row, New York (1965), pp. 82–92.

[22] Serre J.-P., Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures), *Sém. Delange–Poitou–Pisot*, Exposé 19 (1969/1970).

[23] Soulé C., *K-théorie et groupes de Chow de variétés sur un corps fini*, prépublication (1983).

[24] Soulé C., The Rank of Étale Cohomology of Varieties over p-adic or Number Fields, à paraître dans *Compositio Math.* (1983).

[25] Soulé C., On Higher p-adic Regulators, *Lect. Notes in Math.* 854, Springer-Verlag (1980), pp. 372–401.

[26] Tate J., Algebraic Cycles and Poles of Zeta Functions, dans: *Arithmetical Algebraic Geometry*, Harper and Row, New York (1965), pp. 93–110.

[27] Thomason R., *Algebraic K-Theory and Étale Cohomology*, prépublication (1981), 168 pp., et Lettre à E. Friedlander (29/10/82).

[28] Yager R., A Kummer Criterion for Imaginary Quadratic Fields, *Compositio Math.* 47(1)(1982), pp. 31–42.

U.E.R. DE MATHÉMATIQUES, UNIVERSITÉ PARIS VII, TOUR 45-55,
5ÈME ÉTAGE, 2, PLACE JUSSIEU, 75251 PARIS CÉDEX 05, FRANCE

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

RICHARD P. STANLEY*

Combinatorial Applications of the Hard Lefschetz Theorem

1. The hard Lefschetz theorem

Let X be a smooth irreducible complex projective variety of (complex) dimension n (or more generally a Kähler manifold), endowed with the "classical" Hausdorff topology. Let $H^*(X) = H^0(X) \oplus H^1(X) \oplus \ldots \oplus H^{2n}(X)$ denote the singular cohomology ring of X with complex coefficients. (Any field of characteristic zero would do just as well for the coefficient group. In fact, for the most part we could work over $\boldsymbol{Z}$, but this is unnecessary for our purposes.) Since X is projective we may imbed it in some complex projective space P^N. Let H denote a (generic) hyperplane in P^N. Then $X \cap H$ is a closed subvariety of X of real codimension two, and thus by a standard construction in algebraic geometry represents a cohomology class $\omega \in H^2(X)$, called the *class of a hyperplane section.*

THE HARD LEFSCHETZ THEOREM. *Let $0 \leqslant i \leqslant n$. The map $H^i(X) \xrightarrow{\omega^{n-i}} H^{2n-i}(X)$ given by multiplication by ω^{n-i} is an isomorphism of vector spaces.*

This result was first stated by Lefschetz in [18], but his proof was not entirely rigorous. The first complete proof was given by Hodge in [15], using his theory of harmonic integrals. The "standard" proof today uses the representation theory of the Lie algebra $\mathrm{sl}_2(\boldsymbol{C})$ and is due to Chern [4]. Lefschetz' original proof was only recently made rigorous by Deligne (see [22]), who extended it to characteristic p. Other references include [5], [13], [17], [32].

* Partially supported by the National Science Foundation.

2. Unimodality

Since $\omega^{n-i}\colon H^i(X)\to H^{2n-i}(X)$ is bijective for all $0\leqslant i\leqslant n$, it follows that $\omega\colon H^i(X)\to H^{i+1}(X)$ is *injective* for $0\leqslant i\leqslant n-1$ and *surjective* for $n\leqslant i\leqslant 2n-1$. Thus if $\beta_i=\beta_i(X)=\dim H^i(X)$ denotes the i-th Betti number of X, then the sequences $\beta_0,\beta_2,\ldots,\beta_{2n}$ and $\beta_1,\beta_3,\ldots,\beta_{2n-1}$ are *symmetric* and *unimodal*, i.e., $\beta_0\leqslant\beta_2\leqslant\ldots\leqslant\beta_{2[n/2]}$ and $\beta_i=\beta_{2n-i}$, and similarly for $\beta_1,\beta_3,\ldots,\beta_{2n-1}$. This consequence of the hard Lefschetz theorm was well known from the beginning.

There are several examples for which β_i has a combinatorial interpretation. The archetype is the Grassmann variety $X=G_{dn}$ of d-planes in C^n. It was rigorously known since Ehresmann [8] that $\beta_{2i+1}=0$ and that β_{2i} is the number $p(i,d,n-d)$ of partitions of the integer i into $\leqslant d$ parts, with largest part $\leqslant n-d$. The unimodality of the sequence $p(0,d,n-d),p(1,d,n-d),\ldots,p\big(d(n-d),d,n-d\big)$ was first proved by Sylvester [31], and several subsequent proofs have been given. Perhaps the simplest is [28, Cor. 9.6], but no purely combinatorial proof is known. Such a proof would involve an explicit injection from the partitions counted by $p(i,d,n-d)$ to those counted by $p(i+1,d,n-d)$, for $0\leqslant i\leqslant[\frac{1}{2}d(n-d)]$.

More generally, take X to be a generalized flag manifold G/P, where G is a complex semisimple algebraic group and P a parabolic subgroup. The hard Lefschetz theorem then yields the unimodality of the number of elements of length i in the quotient Bruhat order $W^J=W/W_J$, where W is the Weyl group of G and W_J of P ([26, § 3]).

Now let $\mathfrak{g}$ be a complex semisimple Lie algebra and L_λ an irreducible finite-dimensional $\mathfrak{g}$-module with highest weight λ. Let $m_\lambda(\mu)$ denote the multiplicity of the weight μ in L_λ. The *height* $\operatorname{ht}\mu$ of μ is the number of simple roots which need to be added to $-\lambda$ to obtain μ. The polynomial $P_\lambda(q)=\sum_\mu m_\lambda(\mu)q^{\operatorname{ht}\mu}$ was first shown by Dynkin (see [7, p. 332]) to have unimodal coefficients; see also [25]. Recently Lusztig [19] has computed the (complex) intersection cohomology $\boldsymbol{H}^*(\bar{\mathcal{O}}_\lambda)$, as defined by Goresky and Macpherson [10], [11], of certain *Schubert varieties* $\bar{\mathcal{O}}_\lambda$ (in general singular). Namely,

$$\sum_i \dim\boldsymbol{H}^i(\bar{\mathcal{O}}_\lambda)q^i=P_\lambda(q^2).$$

(See [19, Cor. 8.9]). Now the intersection cohomology $\boldsymbol{H}^*(X)$ of any complex projective variety, considered as a module over the singular cohomology

$H^*(X)$, satisfies the hard Lefschetz theorem. Hence Dynkin's result may be regarded as a consequence of the hard Lefschetz theorem for intersection cohomology. It would be interesting to investigate what other sequences of combinatorial interest arise as Betti numbers in intersection cohomology.

3. McMullen's g-conjecture

Let $\mathscr{P}$ be a d-dimensional simplical convex polytope ([14], [21]) with f_i i-dimensional faces, $0 \leqslant i \leqslant d-1$. We call the vector $f(\mathscr{P}) = (f_0, \ldots, f_{d-1})$ the *f-vector* of $\mathscr{P}$. The problem of obtaining information about such vectors goes back to Descartes and Euler. In 1971 McMullen [20], [21, p. 179], drawing on all the available evidence, gave a remarkable condition on a vector $(f_0, \ldots, f_{d-1})$ which he conjectured was equivalent to being an f-vector as above.

To describe this condition, define a new vector $h(\mathscr{P}) = (h_0, \ldots, h_d)$, called the *h-vector* of $\mathscr{P}$, by

$$h_i = \sum_{j=0}^{i} \binom{d-j}{d-i} (-1)^{i-j} f_{j-1},$$

where we set $f_{-1} = 1$. The Dehn–Sommerville equations ([14, §§ 9.2, 9.8], [21, §§ 2.4, 5.1]) assert that $h_i = h_{d-i}$. McMullen's conditions, though he did not realize it, are equivalent to the Dehn–Sommerville equations together with the existence of a graded commutative C-algebra $R = R_0 \oplus \oplus R_1 \oplus \ldots$ where $R_0 = C$, R is generated as a C-algebra by R_1, and $\dim_C R_i = h_i - h_{i-1}$ for $1 \leqslant i \leqslant [d/2]$. In particular, $h_0 \leqslant h_1 \leqslant \ldots \leqslant h_{[d/2]}$, and we are led to suspect the existence of a smooth d-dimensional complex projective variety $X(\mathscr{P})$ for which $\beta_{2i}\big(X(\mathscr{P})\big) = h_i$. If moreover $H^*\big(X(\mathscr{P})\big)$ is generated by $H^2\big(X(\mathscr{P})\big)$, then we can take $R = H^*\big(X(\mathscr{P})\big)/(\omega)$, where (ω) denotes the ideal generated by the class of a hyperplane section, to deduce the necessity of McMullen's conditions. In [6] varieties $X(\mathscr{P})$ are constructed (after some assumptions on $\mathscr{P}$ irrelevant for proving McMullen's conjecture) with all the desired properties except smoothness (despite the misleading statement in [6, Rmk. 3.8]). Although $X(\mathscr{P})$ need not be smooth, its singularities are sufficiently nice that the hard Lefschetz theorem continues to hold [30]. Namely, $X(\mathscr{P})$ is a *V-variety*, i.e., locally it looks like C^n/G where G is a finite group of linear transformations. Thus the necessity of McMullen's condition follows [27]. Sufficiency was proved about the same time by Billera and Lee [1], [2]. For further information see [29]. Some recent work of Kalai [16] suggests that a more elementary proof may be possible, and perhaps an extension to more general objects (such as shellable triangulations of spheres or even triangulations of homology spheres).

4. The Sperner property

Let P be a finite poset (= partially ordered set). We say P is *graded of rank* n if every maximal chain of P has length n (or cardinality $n+1$). We then define the *rank* $\varrho(x)$ of $x \in P$ to be the length l of the longest chain $x_0 < x_1 < \ldots < x_l = x$. Let $P_i = \{x \in P\colon \varrho(x) = i\}$. An *antichain* is a subset A of P of pairwise incomparable elements. Thus each P_i is an antichain. We say that P has the *Sperner property* if no antichain is larger than the largest P_i. This terminology stems from the theorem of E. Sperner [24] that the poset of subsets of an n-element set, ordered by inclusion, has the Sperner property (see also [12]).

If now X is a complex projective variety, we say X has a *cellular decomposition* if there exists a (finite) set $\mathscr{C} = \{C_1, \ldots, C_t\}$ of pairwise disjoint subsets C_i of X, each isomorphic as algebraic varieties to complex affine space C^{n_i}, such that $\cup C_i = X$ and the closure $\bar{C}_i$ (in the classical or Zariski topology) of each C_i is a union of C_j's. We then define a poset $Q^X = Q^X(\mathscr{C})$ to be the set $\mathscr{C}$ ordered by reverse inclusion of the closure of the C_i's. If X is irreducible of dimension n, then Q^X is graded of rank n.

THEOREM 1 ([26, Thm. 2.4]). *If X (as above) is smooth and irreducible, then Q^X has the Sperner property.*

The crucial step in the proof is the use of the hard Lefschetz theorem, and indeed the theorem remains true for any irreducible X (with a cellular decomposition) satisfying the conclusions of the hard Lefschetz theorem. The conclusion to the above theorem is not the strongest possible; see [26] for further details.

The main class of varieties to which the above theorem applies (indeed, the only known class for which the Sperner property of Q^X is non-trivial) are the generalized flag manifolds $X = G/P$ mentioned above. Here the Q^X are the quotient Bruhat orders W^J. For certain choices of G and P the posets Q^X have a special combinatorial significance. In particular, taking $G = \mathrm{SO}(2n+1, C)$ and a certain maximal P, the Spernicity of the poset Q^X can be used to prove the following number-theoretic conjecture of Erdös and Moser [9, eqn. (12)].

THEOREM 2. *Let S be a finite subset of $\boldsymbol{R}$, and for $k \in \boldsymbol{R}$ let $f(S,k)$ denote the number of subsets of S whose elements sum to k. Then for $|S| = 2l+1$, we have*

$$f(S, k) \leqslant f(\{-l, -l+1, \ldots, l\}, 0).$$

For further information in addition to [26], see [23] and the references therein.

It would be interesting to discover other properties of the posets Q^X of Theorem 1 (in addition to the Sperner property). In particular, if Δ denotes the simplicial complex of chains of $\bar{Q}^X$, where $\bar{Q}^X$ denotes Q^X with the bottom and top element removed, then is the geometric realization $|\Delta|$ always a sphere or cell? (It may not even be necessary to assume X is smooth.) This is true when $\dim X \leqslant 2$ or when $X = G/P$. For the latter case see [3].

In [25, Problem 2] it is asked whether some posets arising from irreducible representations of the Lie algebra $\mathrm{sl}(n, \boldsymbol{C})$ have a "symmetric chain decomposition", which is stronger than the Sperner property. In fact, even the Sperner property is open. Perhaps there is a "cellular decomposition for intersection cohomology" of the Schubert varieties $\bar{\mathcal{O}}_\lambda$ discussed in Section 2 which would yield a proof of the Sperner property.

References

[1] Billera L. J. and Lee C. W., Sufficiency of McMullen's Condition for f-Vectors of Simplicial Polytopes, *Bull. Amer. Math. Soc.* **2** (1980), pp. 181–185.

[2] Billera L. J., and Lee C. W., A proof of the Sufficiency of McMullen's Conditions for f-Vectors of Simplicial Convex Polytopes, *J. Combinatorial Theory* (A) **31** (1981), pp. 237–255.

[3] Björner A. and Wachs M., Bruhat Order of Coxeter Groups and Shellability, *Advances in Math.* **43** (1982), pp. 87–100.

[4] Chern S.-S., On a Generalization of Kähler Geometry. In: R. H. Fox, *et al.* (eds.), *Algebraic Geometry and Topology*, Princeton University Press, Princeton, N. J., 1951, pp. 103–124.

[5] Cornalba M., and Griffiths P. A., Some Transcendental Aspects of Algebraic Geometry. In: *Proc. Symp. Pure Math.*, vol. 29, Algebraic Geometry, Arcata (1974), Amer. Math. Soc., Providence, R. I., 1975, pp. 3–110.

[6] Danilov V. I., The Geometry of Toric Varieties, *Russian Math. Surveys* **33** (1978), pp. 97–154; translated from: *Uspekhi Mat. Nauk.* **33** (1978), pp. 85–134.

[7] Dynkin E. B., Maximal Subgroups of the Classical Groups, *AMS Translations*, Ser. 2, vol. 6 (1957), pp. 245–378; translated from: *Trudy Moskov. Mat. Obšč.* **1** (1952), pp. 39–166.

[8] Ehresmann C., Sur la topologie de certains espaces homogènes, *Ann. Math.* **35** (1934), pp. 396–443.

[9] Erdös P., Extremal Problems in Number Theory. In: A. L. Whiteman (ed.), *Theory of Numbers*, Amer. Math. Soc. Providence, R. I., 1965, pp. 181–189.

[10] Goresky M. and Macpherson R., Intersection Homology Theory, *Topology* **19** (1980), pp. 135–162.

[11] Goresky M., and Macpherson R., Intersection Homology Theory II, *Invent. Math.* **72** (1983), pp. 77–129.

[12] Greene C. and Kleitman D. J., Proof Techniques in the Theory of Finite Sets. In: G.-C. Rota (ed.), *Studies in Combinatorics*, Mathematical Association of America, Washington, DC, 1978, pp. 22–79.

[13] Griffiths P. and Harris J., *Principles of Algebraic Geometry*, Wiley, New York, 1978.

[14] Grünbaum B., *Convex Polytopes*, Wiley, New York, NY, 1967.

[15] Hodge W. V. D., *The Theory and Applications of Harmonic Integrals*, 2nd ed., Cambridge University Press, London, 1952.

[16] Kalai G., Characterization of f-Vectors of Families of Convex Sets in R^d, Part I: Necessity of Eckhoff's conditions, *Israel J. Math.*, to appear.

[17] Lamotke K., The Topology of Complex Projective Varieties After S. Lefschetz, *Topology* **20** (1981), pp. 15–51.

[18] Lefschetz S., *L'Analysis situs et la Géométrie Algébrique*, Gauthiers-Villars, Paris, 1924; reprinted in: *Selected Papers*, Chelsea, New York, 1971.

[19] Lusztig G., Singularities, Character Formulas Weight Multiplicities, *Astérisque* **101–102** (1983), pp. 208–229.

[20] McMullen P., The numbers of faces of simplicial polytopes, *Israel J. Math.* **9** (1971), pp. 559–570.

[21] McMullen P. and Shephard G. C., *Convex Polytopes and the Upper Bound Conjecture*, London Math. Soc. Lecture Note Series, vol. 3, Cambridge Univ. Press, London/New York, 1971.

[22] Messing W., Short Sketch of Deligne's Proof of the Hard Lefschetz Theorem. In: *Proc. Symp. Pure Math.*, vol 29, Algebraic Geometry, Arcata (1974), Amer. Math. Soc., Providence, RI, 1975, pp. 563–586.

[23] Proctor R., Representations of sl(2, C) on Posets and the Sperner Property, *SIAM J. Alg. Disc. Math.* **3** (1982), pp. 275–280.

[24] Sperner E., Ein Satz über Untermenge einer endlichen Menge, *Math. Z.* **27** (1928), pp. 544–548.

[25] Stanley R., Unimodal Sequences Arising from Lie Algebras, In: T. V. Narayana *et al.* (eds.), *Combinatorics, Representation Theory and Statistical Methods in Groups*, Dekker, New York, 1980, pp. 127–136.

[26] Stanley R., Weyl Groups, the Hard Lefschetz Theorem, and the Sperner Property, *SIAM J. Alg. Disc. Math.* **1** (1980), pp. 168–184.

[27] Stanley R., The Number of Faces of a Simplicial Convex Polytope, *Advances in Math.* **35** (1980), pp. 236–238.

[28] Stanley R., Some Aspects of Groups Actign on Finite Posets, *J. Combinatorial Theory* (A) **32** (1982), pp. 132–161.

[29] Stanley R., The Number of Faces of Simplicial Polytopes and Spheres, *Ann. N.Y. Acad. Sci.*, to appear.

[30] Steenbrick J. H. M., Mixed Hodge Structure on the Vanishing Cohomology. In:

P. Holm (ed.), *Real and Complex Singularities*, Oslo 1976, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1977, pp. 525–563.
[31] Sylvester J. J., Proof of the Hitherto Undemonstrated Fundamental Theorem of Invariants, *Coll. Math. Papers*, vol. 3, Chelsea, New York, 1973, pp. 117–126.
[32] Weil A., *Introduction à l'étude des variétés Kähleriennes*, Hermann, Paris, 1958.

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA. 02139 U.S.A.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

E. I. ZEL'MANOV

On the Theory of Jordan Algebras

Jordan algebras were introduced in the joint paper of P. Jordan, J. von Neumann and E. Wigner *On the algebraic generalization of the quantum mechanical formalism.* In the usual interpretation of quantum mechanics the observables are hermitian matrices of hermitian operators in the Hilbert space. The set of hermitian matrices is closed with respect to linear combinations $\alpha x+\beta y$ (with real coefficients), but not with respect to the matrix multiplication xy. However, it is closed with respect to the symmetric multiplication $x\circ y = \frac{1}{2}(xy+yx)$. The author's idea has been: (1) to formulate the basic algebraic properties of hermitian matrices in terms of operation $\circ$; (2) to study all algebraic systems which have those properties. Maybe in this way it would be possible to find something which is not exactly the system of hermitian matrices but is similar to it in its algebraic properties.

For the basic properties the author has chosen two identities which are satisfied by $\circ$:

(J1) $x\circ y = y\circ x$ (commutativity),

(J2) $x^2\circ(y\circ x) = (x^2\circ y)\circ x$ (Jordan identity).

Now, linear algebras over the field Φ (of characteristic not equal to 2) which satisfy (J1) and (J2) are called (*linear*) *Jordan algebras.*

Let us consider the most important examples of Jordan algebras [14].

(1) For any associative algebra R let us replace the multiplication xy in R by the symmetric multiplication $x\circ y = \frac{1}{2}(xy+yx)$. The structure we get on R is a Jordan algebra, denoted by $R^{(+)}$.

(2) Let $*\colon R\to R$ be an involution in the associative algebra R, that is, such a linear operator that $(ab)^* = b^*a^*$, $(a^*)^* = a$ for any elements $a, b\in R$. Then the set of $*$-symmetric elements $H(R, *) = \{a\in R \mid a^* = a\}$ is a subalgebra of $R^{(+)}$.

(3) Let $f\colon V\times V\to\Phi$ be a symmetric bilinear form in a vector space V over a field Φ. Define on the direct sum of vector spaces $\Phi\oplus V$ the multiplication $(\alpha\oplus a)(\beta\oplus b)=\big(\alpha\beta+f(a,b)\big)\oplus(\alpha b+\beta a)$. It equips $\Phi\oplus V$ with the structure of Jordan algebra, which is called the *Jordan algebra of a symmetric bilinear form.*

(4) Let C be a Cayley–Dickson algebra over a field Φ with canonical involution $^-\colon C\to C$ (cf. [12], [31]). Consider the 3×3-matrices over C and the involution $*\colon C_3\ni(x_{ij})\to(\bar{x}_{ji})$. The subspace $H(C_3)=\{A\in C_3|\ A^*=A\}$ is a Jordan algebra with respect to the symmetric product. $H(C_3)$ is a simple algebra and $\dim_\Phi H(C_3)=27$. Together with algebras $H(C_3)$ we consider all their forms, that is, those algebras which become isomorphic to algebras of the type $H(C_3)$ under a suitable scalar extension.

We call the algebras from examples (1)–(4) *classical Jordan algebras.* In [14] it was proved that any simple finite-dimensional formally real ($\sum x_i^2=0\Rightarrow x_i=0$) Jordan algebra is classical.

A Jordan algebra is called *special* if it is embeddable into $R^{(+)}$ where R is some associative algebra. Otherwise, it is called *exceptional.* It is easy to see that all the algebras of examples (1)–(3) are special. In 1934 A. Albert [1] proved that the algebras of type (4) are exceptional.

Thus, the formulation and the axiomatic studies of the algebraic properties of the hermitian matrices led us to 4 comparatively small classes, only the 4th of them being essentially new. Here the authors expressed their hope that the pass to the infinite-dimensional case would bring us new types of Jordan algebras. More exactly, it may be formulated as follows:

PROBLEM 1. Do there exist any nonclassical simple Jordan algebras?

A. Albert and N. Jacobson proved that any simple finite-dimensional Jordan algebra is classical. In 1966 N. Jacobson [11] extended this result to the important class of infinite-dimensional Jordan algebras which satisfy the minimality condition for inner ideals. The notions and methods he introduced there had a major influence on the subsequent development of the structure theory. It was shown that inner ideals play the role that one-sided ideals perform in associative algebras. By an *inner ideal* of a Jordan algebra J we mean such a subalgebra B that for any elements $b\in B$, $a\in J$ the element $\{b,a,b\}$ also lies in B; here $\{x,y,z\}=(xy)z+$ $+x(yz)-y(xz)$ is a Jordan triple product. The role of semiprime associative algebras is played by nondegenerate Jordan algebras, that is, those algebras

J which contain no elements $b \neq 0$ with $\{b, J, b\} = 0$. Such elements are called ***absolute zero divisors*** of J.

Using the powerful coordinization theorem, N. Jacobson proved the following analogue of the Wedderbern–Artin theorem: *any nondegenerate Jordan algebra with minimality condition for inner ideals is a finite direct sum of simple ideals which are either classical or Jordan division algebras.*

The smallest ideal M such that J/M is nondegenerate is called the *McCrimmon radical* of J.

In 1977–1978 the author proved that (1) the radical of a Jordan algebra which satisfies either maximality or minimality condition on inner ideals is finite-dimensional and nilpotent [32]; (2) any Jordan division algebra is classical [34], thus finishing the classification of Jordan artinian algebras.

It appears that in the theory of infinite-dimensional associative algebras there were two outstanding problems which essentially stimulated the development of structure methods: (1) the study of algebras with finiteness conditions, and (2) the Burnside-like problems. The same (in even greater degree) is true for the structure theory of infinite-dimensional Jordan algebras.

The Burnside-like problem (or Kurosh problem, cf. [20]) for a class K of algebras is formulated as follows: Is any algebraic algebra $A \in K$ (for any element $a \in A$ there exists such a polynomial $f(x) \in \Phi[x]$ that $f(a) = 0$) locally finite-dimensional?

In 1956 A. I. Shirshov [25] proved by the beautiful combinatorial methods that any special Jordan algebra which satisfies the identity $x^n = 0$ is locally nilpotent. More generally, A. I. Shirshov proved that the Kurosh problem has the positive solution in the class of special Jordan algebras which satisfy an essential polynomial identity (cf. [25], [31]). In this connection he raised the question (cf. [26], [31]).

PROBLEM 2 (A. I. Shirshov). Is any Jordan nil-algebra of bounded degree locally nilpotent?

Later it became clear that this problem is equivalent to the following one:

PROBLEM 2′. Does the McCrimmon radical of a Jordan algebra always lie in its locally nilpotent radical?

and that without the solution of Problems 2, 2′ it would be impossible to solve Problem 1.

1. Connection with Lie algebras

In spite of the 'physical' origin of Jordan algebras they owe the attention of mathematicians mostly due to their startling connections with geometry, analysis, other classes of algebras but first of all with Lie groups and algebras. It was discovered (N. Jacobson, [9]) that the exceptional Lie algebras G_2 may be realized as the differentiation algebras of Cayley–Dickson algebras, algebras F_4 may be realized as the differentiation algebras of Jordan algebras $H(C_3)$ and their forms (M. L. Tomber, [30]). Finally, J. Tits ([28], [29]) used Jordan algebras to construct the models for exceptional algebras E_6, E_7, E_8. Then I. L. Kantor [15] and M. Koecher [16] imbedded an arbitrary Jordan algebra into the $\boldsymbol{Z}$-graded Lie algebra $K(J) = \mathscr{L} = \mathscr{L}_{-1}+\mathscr{L}_0+\mathscr{L}_1$, $\mathscr{L}_i = 0$ for $|i| > 1$.

Conversely, in any $\boldsymbol{Z}$-graded Lie algebra $\mathscr{L} = \mathscr{L}_{-1}+\mathscr{L}_0+\mathscr{L}_i$ the operations $(\mathscr{L}_\sigma, \mathscr{L}_{-\sigma}, \mathscr{L}_\sigma) \ni (a_\sigma, b_{-\sigma}, c_\sigma) \to [[a_\sigma, b_{-\sigma}], c_\sigma]$, $\sigma = \pm 1$ are very close to a Jordan triple product. In particular, if we fix the element $a_{-1} \in L_{-1}$, then the operation $L_1 \times L_1 \ni (x_1, y_1) \to [[x_1, a_{-1}], y_1]$ defines the structure of Jordan algebra on $\mathscr{L}_1$. Thus any Jordan algebra may be treated as $\boldsymbol{Z}$-graded Lie algebra with short $\boldsymbol{Z}$-grading and vice versa.

2. The weakened Burnside problem. A. I. Kostrikin's theorem

The weakened Burnside problem sounds as follows: does there exist such a function $f(n, m)$ of natural arguments $m \geqslant 1, n \geqslant 1$ that the order of any finite group which has n generators and satisfies the identity $x^m = 1$, is $\leqslant f(n, m)$?

For groups of prime exponent $m = p$ this problem was solved in the affirmative in 1958 by A. I. Kostrikin in his fundamental paper [18] (cf. also [19]). At first the group theoretical Burnside problem was reduced to the following problem about Lie algebras: is it true that any Lie algebra of characteristics p which satisfies the Engel identity $[x, y^{p-1}] = [x, y, \ldots \ldots, y] = 0$ is locally nilpotent? Then in the subsequent study of Lie algebras the key role was played by the thin sandwich envelopes (or the 2nd order elements) introduced by A. I. Kostrikin. An element a of a Lie algebra $\mathscr{L}$ is called an *envelope* of the thin sandwich if $[[\mathscr{L}, a], a] = 0$

The following proposition was an important step in the proof of A. I. Kostrikin.

PROPOSITION (A. I. Kostrikin, [18], [19]). *A Lie algebra of characteristics p which is generated by a finite collection of thin sandwich envelopes and satisfies the identity $[x, y^{p-1}] = 0$ is nilpotent.*

3. The Burnside-like problems in Jordan and Lie algebras

Suppose that a Jordan algebra J is generated by its absolute zero divisors. Let us consider the Tits–Kantor–Koecher construction on it, $K(J) = K(J)_{-1}+K(J)_0+K(J)_1$. The Lie algebra $K(J)$ does not a priori satisfy any Engel identity but it is generated by its thin sandwich envelopes.

We have used the techniques of A. I. Kostrikin's paper [18] and the Jordan origin of the Lie algebra $K(J)$ to prove its local nilpotency which is equivalent to the local nilpotency of the algebra J.

This has solved in the affirmative Problem 2′, and hence A. I. Shirshov's Problem 2.

THEOREM 1 ([35], [37]). *The McCrimmon radical of a Jordan algebra lies in its locally nilpotent radical.*

THEOREM 2 ([35], [37]). *Any Jordan nil-algebra of bounded degree is locally nilpotent.*

The Kurosh problem in the class of Jordan PI-algebras has also got the positive solution.

THEOREM 3 ([37]). *Any algebraic Jordan* PI-*algebra is locally finite-dimensional.*

As we have noted before, the connection between Jordan and Lie algebras is two-sided. In fact, Lie algebras with short $\boldsymbol{Z}$-gradings naturally arise when the thin sandwich envelopes or elements with algebraic adjoint operators are involved. So we have used the Jordan techniques to "reverse" the proof of Theorem 1 and to strengthen the above-mentioned proposition of A. I. Kostrikin.

THEOREM 4 ([36]). *A Lie algebra over the ring of scalars $\Phi \ni 1/6$ which is generated by a finite collection of thin sandwich envelopes is nilpotent.*

As usual, by $\mathrm{ad}(a)$, $a \in \mathscr{L}$ we denote the commutation operator with an element a; $\mathrm{ad}(a)\colon \mathscr{L} \ni x \to [x, a]$. We say that $\mathscr{L}$ is an *algebra with algebraic adjoint representation* if for any element $a \in \mathscr{L}$ there exists such a polynomial $f(x) \in \Phi[x]$ that $f\big(\mathrm{ad}(a)\big) = 0$.

The Jordan techniques (and Theorem 4) proved to be useful for the solution of one more Burnside-like problem for Lie algebras.

THEOREM 5 ([40]). *A Lie algebra with algebraic adjoint representation over the field of characteristic* 0 *which satisfies the polynomial identity is locally finite-dimensional.*

Remark that A. N. Grishkov [7] used Theorem 4 to prove that in a Lie algebra of characteristic zero the thin sandwich envelopes generate the locally nilpotent ideal.

4. The classification theorems for Jordan algebras

After a positive solution of Problems 2 and $2'$, it has become possible to solve in the affirmative Problem 1.

THEOREM 6 ([39]). *Any simple Jordan algebra is isomorphic to one of the following algebras:* (1) $R^{(+)}$ *where* R *is a simple associative algebra;* (2) $H(R, *)$ *where* R *is a simple associative algebra with the involution* $*$: $R \to R$; (3) *the Jordan algebra of a nondegenerate symmetric bilinear form in a vector space* V *over some extension* Γ *of the basic field* Φ, $\dim_\Gamma V > 1$; (4) *the simple exceptional Jordan algebra which is* 27*-dimensional over its centre.*

Conversely, if R is a simple associative algebra, then by I. N. Herstein's theorems (cf. [8]) Jordan algebras of types (1) and (2) are simple.

We call an algebra *prime* if the product of any two of its nonzero ideals is nonzero. Any nondegenerate Jordan algebra may be approximated by prime nondegenerate Jordan algebras. That is why from the ring-theoretical point of view it would be useful to determine not only simple algebras but all prime nondegenerate Jordan algebras. By "to determine" we mean to find out to what extent they are close to the classical Jordan algebras. The following two theorems show that they are fairly close.

A prime Jordan algebra J is said to be an *Albert ring* if its centre $Z(J)$ is nonzero and its central closure $Z(J)^{-1}J$ is a simple exceptional algebra of dimension 27 over its centre $Z(J)^{-1}Z(J)$.

THEOREM 7 ([39]). *A prime nondegenerate Jordan algebra is either special or an Albert ring.*

THEOREM 8 ([39]). *Let* J *be a special prime nondegenerate Jordan algebra Then one of the following assertions is valid:*

(I) *the centre* $Z(J)$ *is nonzero and the central closure* $Z(J)^{-1}J$ *is a Jordan algebra of nondegenerate symmetric bilinear form over the field* $Z(J)^{-1}Z(J)$.

(II) J *contains the nonzero ideal* I *which is invariant under all automorphisms and all differentiations of* J *and either*

(II.1) $I \simeq R^{(+)}$, R *is a prime associative algebra and* $R^{(+)} \trianglelefteq J \subseteq Q(R)^{(+)}$ *where* $Q(R)$ *is a Martindale quotient ring of* R (cf. [22]), *or*

(II.2) $I \simeq H(R, *)$, *R being a prime associative algebra with involution and* $H(R, *) \trianglelefteq J \subseteq H(Q(R), *)$.

Conversely, algebras of types (I), (II) *are prime and nondegenerate.*

5. Lie algebras with finite grading

Lie algebras A_n, B_n, C_n, D_n and also the exceptional algebras E_6, E_7 are equipped with the nontrivial short $\mathbf{Z}$-grading $\mathscr{L} = \mathscr{L}_{-1}+\mathscr{L}_0+\mathscr{L}_1$. Lie algebras with a short grading may be constructed also from an arbitrary Jordan algebra (Jordan pair, Jordan triple system, cf. [21], [24]). Algebras G_2, F_4, E_8 do not have a short grading but they have a slightly longer $\mathbf{Z}$-grading $\mathscr{L} = \mathscr{L}_{-2}+\mathscr{L}_{-1}+\mathscr{L}_0+\mathscr{L}_1+\mathscr{L}_2$. For the construction of these algebras the so-called *J-ternary algebras* were introduced and studied (cf. [2], [5]). However, in certain considerations connected with infinite-dimensional Lie algebras or finite-dimensional Lie algebras over nonclosed fields there occur $\mathbf{Z}$-gradings $\mathscr{L} = \sum_{i=-n}^{n} \mathscr{L}_i$ of arbitrary finite length. In fact, under certain restrictions on the characteristics of the basic field, the theory of Lie algebras with finite grading has turned out to be "parallel" to the Jordan theory. This parallelism is not formal: the most important notions and methods of the theory of Jordan algebras admit natural analogues for Lie algebras with finite grading. That is why Lie algebras with finite grading may be by right included in the Jordan theory as its most general object (so far). This ideology has made it possible to classify the simple (infinite-dimensional) Lie algebras with finite grading.

THEOREM 9 ([38]). *Let* $\mathscr{L} = \sum_{i=-n}^{n} \mathscr{L}_i$ *be a simple $\mathbf{Z}$-graded Lie algebra over a field of characteristic* $p \geqslant 4n+1$ (*or* 0), $\sum_{i\neq 0} \mathscr{L}_i \neq 0$. *Then one of the following assertions is valid*:

(1) *there exists a simple $\mathbf{Z}$-graded associative algebra* $R = \sum_{i=-n}^{n} R_i$ *such that* $\mathscr{L} \simeq [R^{(-)}, R^{(-)}]/Z$, *where Z is the centre of the commutant* $[R^{(-)}, R^{(-)}]$,

(2) *there exists a simple $\mathbf{Z}$-graded associative algebra* $R = \sum_{i=-n}^{n} R_i$ *with involution* $*: R\to R$, $R_i^* = R_i$, *such that* $\mathscr{L} \simeq [S, S]/Z([S, S])\cap R_0$ (*where* $S = S(R, *) = \{a \in R \mid a^* = -a\}$) *is a Lie algebra of $*$-skew-symmetric elements*,

(3) $\mathscr{L}$ *is isomorphic to the Tits–Kantor–Koecher construction of the Jordan algebra of a symmetric bilinear form over some extension of the basic field*,

(4) $\mathscr{L}$ *is of one of the types* G_2, F_4, E_6, E_7, E_8, D_4.

The isomorphisms in (1) and (2) are isomorphisms of graded algebras. In the proof of the theorem the cases $n = 1$ and $n > 1$ should be treated separately. However, the more difficult case $n = 1$ was considered earlier in terms of the Jordan pairs [37].

One may consider an even more general situation: Λ is a torsion free abelian group, $\mathscr{L} = \sum_{\lambda\in\Lambda} \mathscr{L}_\lambda$, a Λ-graded Lie algebra, the set $\{\lambda | \mathscr{L}_\lambda \neq 0\}$ is finite. Then under certain restrictions on the characteristic of the basic field the analogue of Theorem 9 is valid.

For other aspects and results in Jordan systems and their relations see monographs [4], [8], [12], [13], [17], [21], [24], [27], [31] and surveys [3], [23], [26].

References

[1] Albert A. A., On a certain algebra of quantum mechanics, *Ann. Math.* **36** (1934) pp. 65–73.

[2] Allison B. N., A construction of Lie algebras from J-ternary algebras, *Amer. J. Math.* **98** (1976), pp. 285–294.

[3] Bakhturin Yu. A., Slin'ko A. M., and Shestakov I. P., Nonassociative rings, *Itogi nauki i tekhniki, series: algebra, topology, geometry* **18**, pp. 3–72, Moscow, 1981.

[4] Braun H., Koecher M., *Jordan Algebren*, Springer Verlag, Berlin, 1966.

[5] Faulkner J. and Ferrar J., On the structure of symplectic ternary algebras, *Indag. Math.* **34** (1972), pp. 247–265.

[6] Golod E. S., On nil-algebras and finitely approximated groups, *Izv. AN SSSR* **28** (1964), pp. 273–276.

[7] Grishkov A. N., On the local nilpotency of the Lie algebra ideals generated by the 2nd order elements, *Siberian Math. J.* **23** (1982), pp. 181–182.

[8] Herstein I. N., *Rings with involutions*, Chicago lectures in Math., Chicago, 1976.

[9] Jacobson N., Cayley numbers and normal Lie algebras of type G, *Duke Math. J.* **5** (1939), pp. 775–783.

[10] Jacobson N., *Exceptional Lie algebras*, Mimeographed notes, Yale Univ., New Haven, 1958.

[11] Jacobson N., Structure theory for a class of Jordan algebras, *Proc. Nat. Acad. Sci. USA* **55** (1966), pp. 243–251.

[12] Jacobson N., *Structure and representations of Jordan algebras*, AMS Colloq. Publ. v. 39, Providence, R. I., 1969.

[13] Jacobson N., *Structure theory of Jordan algebras*, Lecture Notes, University of Arkansas, Feyetteville, 1981.

[14] Jordan P., von Neumann J., and Wigner E., On an algebraic generalization of the quantum mechanical formalism, *Ann. Math.* **36** (1934), pp. 29–64.

[15] Kantor I. L., Some generalizations of Jordan algebras, *Trudy Sem. Vector. Tensor. Anal.* **16** (1972), pp. 407–499.

[16] Koecher M., Imbedding of Jordan algebras into Lie algebras, I., *Amer. J. Math.* **89** (1967), pp. 787–816.

[17] Koecher M., *An elementary approach to bounded symmetric domains*, Lecture Notes, Rice University, Houston, 1969.
[18] Kostrikin A. I., On the Burnside problem, *Izv. AN SSSR* **23** (1959), pp. 3–34.
[19] Kostrikin A. I., Sandwiches in Lie algebras, *Matem. Sb.* **110** (1979), pp. 3–12.
[20] Kurosh A. G., Problems in ring theory which are related to the Burnside problem on periodic groups, *Izv. AN SSSR* **5** (1941), pp. 233–240.
[21] Loos O., *Jordan Pairs*, Lecture Notes in Math., v. **460**, Springer-Verlag, New York, 1975.
[22] Martindale W. S., III, Prime rings with involution and generalized polynomial identities, *J. Algebra* **22** (1972), pp. 502–516.
[23] McCrimmon K., Jordan algebras and their applications, *Bull. AMS* **84** (1978), pp. 612–627.
[24] Meyberg K., *Lectures on algebras and triple systems*, Lecture Notes, University of Virginia, Charlottesville, 1972.
[25] Shirshov A. I., On certain nonassociative nil-rings and algebraic algebras, *Mat. Sb.* **41** (1957), pp. 381–394.
[26] Shirshov A. I., Some problems in the theory of rings which are nearly associative, *Uspekhi mat. nauk.* **13** (1958), pp. 3–20.
[27] Springer T. A., *Jordan algebras and algebraic groups*, Springer-Verlag, New York, 1973.
[28] Tits J., Une Classe d'algèbres de Lie en relation aves les algèbres de Jordan, *Indag. Math.* **24** (1962), pp. 530–535.
[29] Tits J., Algèbres alternatives, algèbres de Jordan et algèbres de Lie exceptionelles. I. Construction, *Indag. Math.* **28** (1966), pp. 223–237.
[30] Tomber M. L., Lie algebras of type *F*, *Proc. AMS* **4** (1953), pp. 759–768.
[31] Ževlakov K. A., Slin'ko A. M., Shestakov I. P., and Shirshov A. I., *Rings which are nearly associative*, Nauka, Moscow, 1978.
[32] Zel'manov E. I., Jordan algebras with finiteness conditions, *Algebra i logika* **17** (1978), pp. 693–704.
[33] Zel'manov E. I., On prime Jordan algebras, *Algebra i logika* **18** (1979), pp. 162–175.
[34] Zel'manov E. I., Jordan division algebras, *Algebra i logika* **18** (1979), pp. 286–310.
[35] Zel'manov E. I., Jordan nil-algebras of bounded degree, *DAN* **249** (1979), pp. 30–33.
[36] Zel'manov E. I., Absolute zero divisors in Jordan pairs and Lie algebras, *Mat. Sb.* **117** (1980), pp. 611–629.
[37] Zel'manov E. I., Absolute zero divisors and algebraic Jordan algebras, *Siberian Math. J.* **23** (1982), pp. 100–116.
[38] Zel'manov E. I., Classification theorems for Jordan systems, *Siberian Math. J.* **23** (1982), pp. 186–187.
[39] Zel'manov E. I., Lie algebras with finite grading, *Materials of the 5th All-Union Ring Th. Conference*, Novosibirsk, 1982, p. 58.
[40] Zel'manov E. I., On prime Jordan algebras. II, *Siberian Math. J.* **24** (1983), pp. 89–104.
[41] Zel'manov E. I., Algebraic Lie algebras, *Materials of the 4th All-Union Ring Th. Conference, Kishinev*, 1980, p. 39.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

A. N. ANDRIANOV

Integral Representations of Quadratic Forms by Quadratic Forms: Multiplicative Properties

1. Introduction

Modular forms with respect to congruence subgroups of the integral symplectic group $\Gamma^n = \mathrm{Sp}_n(\boldsymbol{Z})$ were introduced as a generalization of theta-series of integral positive quadratic forms and have allowed us to clarify a number of their properties. The corner stone of the magnificent progress of the theory of modular forms of one variable during last two decades was laid by E. Hecke by his discovery that all Fourier coefficients of a modular form can be expressed through eigenvalues of a multiplicative family of invariantly defined linear operators. If $n > 1$, it seems that the Hecke operators are insufficient to provide a similar result, although certain relations between their eigenvalues and the Fourier coefficients of eigenfunctions were obtained (see § 3 below). Hence the situation in the general case looks worse than the classical one. However, a modular optimist can argue that the opinion is due to a too narrow approach, which is based exclusively on the ideology of the one variable case. It seems there is no other way to find an adequate approach but to go back to perhaps the only observable kind of modular forms of many variables, namely to the theta-series and to look what Hecke operators can do with them and with their Fourier coefficients. In this report we are going to present some results, observations and conjectures in this direction.

2. Modular forms, theta-series, Hecke operators

A modular form of order n, integral weight k and Dirichlet character χ modulo q with respect to the congruence subgroup

$$\Gamma_0^n(q) = \left\{M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Gamma^n;\ C \equiv 0 \pmod{q}\right\}$$

of the group $\Gamma^n = \mathrm{Sp}_n(\boldsymbol{Z})$ is a function $F(Z)$ of $n(n+1)/2$ complex variables which is holomorphic on the Siegel upper halfplane of order n

$$H_n = \{Z = X+iY \in M_n(\boldsymbol{C});\ {}^tZ = Z,\ Y > 0\},$$

which for each matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Gamma_0^n(q)$ satisfies the functional equation

$$F\big((AZ+B)(CZ+D)^{-1}\big) = \chi(\det D)\det(CZ+D)^k F(Z)$$

and which, if $n = 1$, is holomorphic at all vertices of the group $\Gamma_0^1(q)$. All such functions form a linear space $\mathfrak{M}_k^n(q, \chi)$ over $\boldsymbol{C}$. A modular form $F \in \mathfrak{M}_k^n(q, \chi)$ can be expanded in a Fourier series of the form

$$F(Z) = \sum_{R \in \mathfrak{N}_n} f(R)\exp\big(\pi i\sigma(RZ)\big),$$

where $\mathfrak{N}_n$ is the set of all symmetric, integral and semidefinite matrices of order n with even diagonal elements and $\sigma(RZ)$ is the trace of the matrix RZ. The Fourier coefficients $f(R)$ satisfy the relations

$$f(UR{}^tU) = \chi(\det U)(\det U)^k f(R) \qquad \big(U \in \mathrm{GL}_n(\boldsymbol{Z})\big).$$

For each matrix $Q \in \mathfrak{N}_m^+ = \{R \in \mathfrak{N}_m;\ R > 0\}$ and $n = 1, 2, \ldots$ the theta-series of Q of order n is defined by

$$\begin{aligned}\theta^n(Z, Q) &= \sum_{M \in M_{m,n}(\boldsymbol{Z})} \exp\big(\pi i\sigma({}^tMQMZ)\big) \\ &= \sum_{R \in \mathfrak{N}_n} r(Q, R)\exp\big(\pi i\sigma(RZ)\big),\end{aligned}$$

where $Z \in H_n$ and $r(Q, R)$ is the number of solutions in integral $(m \times n)$-matrices X of the equation ${}^tXQX = R$, i.e., the number of integral representation of the quadratic form with matrix R by the form with matrix Q. It is proved in [1] that for even $m = 2k$ we have

$$\theta^n(Z, Q) \in \mathfrak{M}_k^n(q, \chi_Q),$$

where q is the level of Q and χ_Q is the character of Q.

Let $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be an integral $(2n \times 2n)$-matrix which satisfies the conditions

$${}^tM\begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix}M = \delta(M)\begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix}, \qquad \delta(M) > 0, \quad \big(\delta(M), q\big) = 1,$$
$$C \equiv 0 \,(\mathrm{mod}\, q),$$

where $E = E_n$ is the unit matrix of order n. Then the operator

$$T(M)\colon F \to F|_{k,\chi}T(M)$$
$$= \delta(M)^{nk-n(n+1)/2} \sum \chi(\det A')\det(C'Z+D')^{-k}F\big((A'Z+B')(C'Z+D')^{-1}\big),$$

where the sum is taken over

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \in \Gamma_0^n(q)\backslash \Gamma_0^n(q)M\Gamma_0^n(q),$$

maps the space $\mathfrak{M}_k^n(q, \chi)$ into itself. Finite linear combinations of operators $T(M)$ with constant coefficients are called the *Hecke operators*. A Hecke operator $\sum_i a_i T(M_i)$ is called a *square-operator* (resp. a p-operator, where p is a prime) if $\delta(M_i)$ are squares (resp. are of the form p^{d_i}). We also define

$$T(l) = \sum_{(\Gamma_0^n(q)M\Gamma_0^n(q)),\delta(M)=l} T(M), \quad \text{where} \quad (l, q) = 1.$$

All Hecke operators on $\mathfrak{M}_k^n(q, \chi)$ (resp. all square-operators, all p-operators) form a commutative ring $L = L(n, k, q, \chi)$ (resp. L_{sq}, L_p).

A space $\mathfrak{M}_k^n(q, \chi)$ is spanned by eigenforms [2], i.e. by the forms $F \in \mathfrak{M}_k^n(q, \chi)$ which satisfy

$$F|_{k,\chi}X = \lambda_F(X)F \quad \text{for all } X \in L.$$

If $F \in \mathfrak{M}_k^n(q, \chi)$ is an eigenform and p is a prime, the map $X \to \lambda_F(X)$ gives us a representation of L_p into $\boldsymbol{C}$. The representation can be obtained in the following manner [3]. For each M of the form described above with $\delta(M) = p^d$ all representatives M' from the different left cosets $\Gamma_0^n(q)M' \subset \Gamma_0^n(q)M\Gamma_0^n(q)$ can be chosen in a triangular form:

$$M' = \begin{bmatrix} p^{d} (D')^{-1} & B' \\ 0 & D' \end{bmatrix}, \quad \text{where} \quad D' = \begin{bmatrix} p^{d_1'} & * & \dots & * \\ 0 & p^{d_2'} & \dots & * \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p^{d_n'} \end{bmatrix}.$$

Then there exist non-zero complex numbers $a_i = a_i(p, F)$ $(i = 0, 1, \dots, n)$ such that

$$\lambda_F\big(T(M)\big) = a_0^d \sum_{M'} \prod_{j=1}^{h} (a_j p^{-j})^{d_j'}, \quad (T(M) \in L_p).$$

The numbers $a_0, a_1, \dots, a_n$, which we shall call the *p-parameters* of the eigenform F, are determined by λ_F and p, uniquely up to the action of the Weyl group of Sp_n. They satisfy

$$a_0^2 a_1 \dots a_n = p^{nk-n(n+1)/2}\chi(p^n).$$

3. Fourier coefficients of eigenforms and eigenvalues

Let $F \in \mathfrak{M}_k^n(q, \chi)$ be an eigenform with the Fourier coefficients $f(R)$ $(R \in \mathfrak{N}_n)$ and the eigenvalues $\lambda_F(X)$ $(X \in L)$. Since the eigenvalues $\lambda_F(X)$ have certain multiplicative properties, it is natural to try to express $f(R)$ through $\lambda_F(X)$. If $n = 1$, this was done by Hecke. If $n > 1$, only some relations between the Fourier coefficients and the eigenvalues were obtained. The only type of relations which was found for all n is given by

THEOREM 1 [4]. *Let $F \in \mathfrak{M}_k^n(q, \chi)$ be an eigenform with the Fourier coefficients $f(R)$ $(R \in \mathfrak{N}_n)$. Then for any fixed matrix $R \in \mathfrak{N}_n^+$ and any completely multiplicative function ψ on positive integers which grows not faster than a polynomial and satisfies the condition*

$$\psi(p) = 0 \quad \text{if } p \text{ is a prime, } p | q \det R,$$

the following identity is valid, provided the real part of the complex variable s is sufficiently large:

$$\sum_{M \in \mathrm{SL}_n(\mathbf{Z}) \setminus M_n^+(\mathbf{Z})} \psi(\det M) f(MR\,{}^tM)(\det M)^{1-k-s}$$

$$= f(R)\Big\{L(s+n/2, \psi\chi_R\chi) \prod_{i=0}^{n/2-1} L(2s+2i, \psi^2\chi^2)\Big\}^{-1} D_F(s, \psi)$$

if n is even, and

$$= f(R)\Big\{\prod_{i=0}^{(n-1)/2} L(2s+2i, \psi^2\chi^2)\Big\}^{-1} D_F(s, \psi)$$

f n is odd, where $M_n^+(\mathbf{Z})$ is the set if all integral matrices of order n with positive determinants,

$$L(t, \omega) = \sum_{n=1}^{\infty} \frac{\omega(n)}{n^t},$$

χ_R is the character of R and

$$D_F(s, \psi) = \prod_p \Big\{\Big(1 - \frac{\psi(p)\chi(p)}{p^s}\Big) \prod_{i=1}^{n} \Big(1 - \frac{\psi(p)\chi(p)}{p^s}\, a_i^{-1}(p, F)\Big) \times$$

$$\times \Big(1 - \frac{\psi(p)\chi(p)}{p^s}\, a_i(p, F)\Big)\Big\}^{-1}$$

where p runs through all prime numbers and for each prime p with $(p, q) = 1$, $\alpha_i(p, F)$ are the p-parameters of F.

In [5] the restriction $\psi(p) = 0$ if p divides $q \det R$ was replaced by $\psi(p) = 0$ if p divides q. In this case the factor $f(R)$ on the right-hand side of the identity should be replaced by a finite sum. Relations of a different kind were obtained earlier in [6], [7] but, unfortunately, only for $n = 2$. All attempts to extend them have failed.

The identities of Theorem 1 not only give the multiplicativity of certain linear combinations of the Fourier coefficients of an eigenform F but also allow us to investigate some analytical properties of the associated Euler product $D_F(s, \psi)$, where ψ is a Dirichlet character: The Dirichlet series on the left-hand side of the identities can be obtained from F by means of an integral transformation. It allows us to prove that $D_F(s, \psi)$ has an analytical continuation over the whole s-plane and in some cases satisfies a functional equation [8].

4. Action of Hecke operators on theta-series

The first question on the way back from general modular forms to theta-series was whether Hecke operators transform theta-series again into linear combinations of theta-series. A positive answer was obtained in [9] for all square-operators and, in some cases, for all Hecke operators. More specifically, the ring L of all Hecke operators on $\mathfrak{M}_k^n(q, \chi)$ is generated by the square-subring L_{sq} and by the operators $T(p)$ for all prime p such that $(p, q) = 1$. Let $Q \in \mathfrak{R}_{2k}^+$ and $\theta^n(Z, Q) \in \mathfrak{M}_k^n(q, \chi)$. It was proved that the form $\theta^n(Z, Q)|_{k,\chi} T(p)$ is a linear combination of theta-series if $\chi_Q(p) = 1$ or $\chi_Q(p) = -1$ and $n \geqslant k$. It is still unknown whether this is true if $\chi_Q(p) = -1$ and $1 \leqslant n < k$.

In all cases mentioned above it was proved that for $Q \in \mathfrak{R}_{2k}^+$ and $\theta^n(Z, Q) \in \mathfrak{M}_k^n(q, \chi)$ we have

$$\theta^n(Z, Q)|_{k,\chi} X = \sum_i \alpha(Q, Q_i; X)\, \theta^n(Z, Q_i)$$

where Q_i runs through a system of representatives of all classes $\{Q'\} = \{{}^tUQ'U;\ U \in \mathrm{GL}_{2k}(\boldsymbol{Z})\}$ with $\det Q' = \det Q$. At first no formulae for the coefficients $\alpha(Q, Q_i; X)$ were obtained. Such formulae were found in [5]. Another approach was proposed later in [10]. Although explicit, the formulae have not clarified the arithmetical meaning of the coefficients. The clarification was made, in some cases, in [11] and [12]. We shall state here the corresponding results only in the case $n = k$.

THEOREM 2. *Let $R_m(D)$ be a system of representatives of all primitive classes $\{Q\} \in \mathfrak{N}_m^+$ with* $\det Q = D$ *(primitivity means that $d^{-1}Q \in \mathfrak{N}_m$ for $d \in \boldsymbol{Z}$ implies $d = 1$). Suppose that $m = 2k$, p is a prime and $(p, D) = 1$. Then*

$$\theta^k(Z, Q)|_{k,\chi}T(p) = \sum_{Q' \in R_m(D)} e(Q')^{-1} r(Q, pQ')\, \theta^k(Z, Q'), \tag{1}$$

where $Q \in R_m(D)$, $\chi = \chi_Q = \chi_D$, $r(Q, Q_1)$ is the number of integral representations of the quadratic form Q_1 by Q and

$$e(Q') = r(Q', Q').$$

Suppose that $m = 2$, l is an integer and $(l, D) = 1$. Then

$$\theta^1(Z, Q)|_{1,\chi}T(l) = \sum_{Q' \in R} e(Q')^{-1} r(Q, lQ')\, \theta^1(Z, Q'). \tag{2}$$

If would be interesting to extend formula (1) to a more general type of Hecke operators.

5. Relations between numbers of integral representations

The fact that the numbers of integral representations of quadratic forms by the forms of the same order appear as elements of matrices of representation of Hecke operators on the spaces of theta-series allows us in principle to find for them some multiplicative relations of the same type as for the corresponding operators. For example, if $m = 2$, it follows from formula (2) and from the classical formulae for the products $T(l)T(l_1)$ that

$$E_D(l)E_D(l_1) = \sum_{d|l, l_1} \chi_D(d) E_D(ll_1/d^2),$$

where $D > 0$, $(ll_1, D) = 1$, $E_D(l) = \left(e(Q_i)^{-1} r(Q_i, lQ_j)\right)$, $Q_1, \ldots, Q_H$ is some ordering of $R_2(D)$ and $\chi_D = \chi_{Q_1}$ [12].

But the formulae can give more than that. Consider, for example, formula (1). By direct computation one can find the relation

$$\theta^k(Z, Q)|_{k,\chi}T(p) = \sum_{R \in \mathfrak{N}_k} r_p(Q, R) \exp\left(\pi i \sigma(RZ)\right),$$

where

$$r_p(Q, R) = \sum_{i=0}^{k} \chi(p^i) p^{(i^2-i)/2} \sum_{D \in \mathrm{SL}_k(\boldsymbol{Z}) \backslash (D_{k-i}), DR\, {}^tD \in p\mathfrak{N}_k} r(Q, p^{-1}DR^+D),$$

$$D_j = \operatorname{diag}(\underbrace{1, \ldots, 1}_{k-j}, \underbrace{p, \ldots, p}_{j}), \qquad (D_j) = \mathrm{SL}_k(\boldsymbol{Z}) D_j \mathrm{SL}_k(\boldsymbol{Z}).$$

Comparing the Fourier coefficients of the series on both sides of (1), we get the relations

$$r_p(Q_i, R) = \sum_{j=1}^{H} e(Q_j)^{-1} r(Q_i, pQ_j) r(Q_j, R), \tag{3}$$

where $R \in \mathfrak{N}_k$, $(Q_1, \ldots, Q_H) = R_{2k}(D)$, $D > 0$. Let us assume that the following conjecture is true.

CONJECTURE 1. *The theta-series*

$$\theta^k(Z, Q_1), \ldots, \theta^k(Z, Q_H)$$

of order k, where $(Q_1, \ldots, Q_H) = R_{2k}(D)$, k, $D > 0$, are linearly independent.

Then there exist matrices $R_1, \ldots, R_H \in \mathfrak{N}_k^+$ such that

$$\det\big(r(Q_j, R_t)\big) \neq 0.$$

Therefore it follows from (3) that the following matrix relations are true:

$$\big(e(Q_1)^{-1} r(Q_i, pQ_1), \ldots, e(Q_H)^{-1} r(Q_i, pQ_H)\big)$$
$$= \big(r(Q_j, R_t)\big)^{-1} \big(r_p(Q_i, R_1), \ldots, r_p(Q_i, R_H)\big),$$

where $i = 1, 2, \ldots, H$. The relations express the numbers of integral representation of certain matrices of order $2k$ by matrices of order $2k$ through the numbers of integral representations of some matrices of order k by the matrices of order $2k$. One can hope that there exist more relations of this kind. If $k = 1$, more explicit relations were obtained unconditionally in [12].

6. Linear independence of theta-series and degenerations of Hecke operators

Conjecture 1 is true if $k = 1$ (it follows, for example, from the fact that each ideal class of a subring of the ring of integers of the imaginary quadratic field $Q\big(\sqrt{-D}\big)$ contains infinitely many prime ideals). If $k > 1$, it has a number of numerical and theoretical supports. Some of them are provided by the theory of Hecke operators. For example, the following statement on degenerations of the action of Hecke operators on theta-series is a direct consequence of Conjecture 1:

THEOREM 3. *Let $F \in \mathfrak{M}_k^n(q, \lambda)$, where $1 \leqslant k \leqslant n \leqslant 2k$, be an eigenform which is a linear combination of theta-series $\theta^n(Z, Q) \in \mathfrak{M}_k^n(q, \chi)$ with $Q \in {}^{+}_{2k}$. Then for each prime p which does not divide q, $n-k$ of the p-parameters*

$a_1(p, F), \ldots, a_n(p, F)$ *of* F *have fixed values:*

$$a_i(p, F) = p^{-i}\chi(p) \qquad (i = 1, \ldots, n-k).$$

On the other hand, the theorem was proved unconditionally in [13] in virtue of explicit formulae for the action of Hecke operators on theta-series.

7. Duality in Siegel's theorem and the genus operator

It is essential for applications of Hecke operators when other invariant operators commute with Hecke operators. For example, the role of Jarkovskaya's commutation relations for the Siegel operator Φ: $\mathfrak{M}_k^n(q, \chi) \to \mathfrak{M}_k^{n-1}(q, \chi)$ is well known. A hypothetical operator of the kind emerges from a duality in Siegel's theorem on the mean number of integral representations of a quadratic form by a genus of another form. The theorem gives an expression for the sums of the form

$$\sigma_1(Q, R) = \mu(Q)^{-1} \sum_{i=1}^{H} e(Q_i)^{-1} r(Q_i, R),$$

where $Q \in \mathfrak{R}_m^+$, $R \in \mathfrak{R}_n^+$, $Q_1, \ldots, Q_H$ is a system of representatives of al classes in the genus $[Q]$ of Q and

$$\mu(Q) = \sum_{i=1}^{H} e(Q_i)^{-1}$$

is the weight of the genus $[Q]$, through the numbers of solutions of the corresponding congruences. It is only natural to ask whether it is possible to say something about the dual sums

$$\sigma_2(Q, R) = \mu(R)^{-1} \sum_{j=1}^{h} e(R_j)^{-1} r(Q, R_j),$$

where $R_1, \ldots, R_h$ is a system of representatives of all classes in the genus $[R]$ of R and $\mu(R)$ is the weight of the genus $[R]$. It is proved in [14] that

$$\sigma_1(Q, R) = \sigma_2(Q, R) \quad \text{if} \quad m = n. \tag{4}$$

Relations (4) can be formulated in another form. We shall say that the genus operator $F \to [F]$ is defined for a modular form

$$F(Z) = \sum_{R \in \mathfrak{R}_n} f(R) \exp\left(\pi i \sigma(RZ)\right) \in \mathfrak{M}_k^n(q, \chi),$$

where $n \leqslant 2k$ if there exists a modular form

$$[F](Z) = \sum_{R \in \mathfrak{N}_n} [f](R) \exp\big(\pi i \sigma(RZ)\big) \in \mathfrak{M}_k^n(q, \chi)$$

such that

$$[f](R) = \mu(R)^{-1} \sum_{j=1}^{h(R)} e(R_j)^{-1} f(R_j) \qquad \text{for all } R \in \mathfrak{N}_n^+,$$

where $R_1, \ldots, R_h$ is a system of representatives of all classes in the genus $[R]$ and $\mu(R)$ is the weight of the genus. Since the space $\mathfrak{M}_k^n(q, \chi)$ does not contain singular forms, provided that $n \leqslant 2k$, the form $[F]$ is uniquely defined if the genus operator $F \to [F]$ is defined. Relations (4) for even $m = 2k$ mean that the genus operator is defined for each theta-series $F(Z) = \theta^{2k}(Z, Q)$, where $Q \in R_{2k}^+$, and has as the image

$$[F](Z) = \mu(Q)^{-1} \sum_{i=1}^{H} e(Q_i)^{-1} \theta^{2k}(Z, Q_i)$$

(the notation is as above).

Moving towards a generalization of the relations (4) to the cases where $m \geqslant n$, one can propose the following

CONJECTURE 2. *The genus operator is defined for each theta-series $\theta^n(Z, Q)$, where $Q \in \mathfrak{N}_{2k}^+$ and $n \leqslant 2k$.*

More generally one can propose

CONJECTURE 3. *The genus operator is defined for each modular form $F \in \mathfrak{M}_k^n(q, \chi)$, where $1 \leqslant n \leqslant 2k$, $q \geqslant 1$, and χ is the Dirichlet character module q, $\chi(-1) = (-1)^k$.*

To start the study of the genus operator, the following theorem is proved in [14]:

THEOREM 4. *Suppose that the genus operator is defined for a modular form $F \to \mathfrak{M}_k^n(q, \chi)$, where $n \leqslant 2k$, and X is a Hecke operator on $\mathfrak{M}_k^n(q, \chi)$. Then the genus operator is defined for the form $F|_{k,\chi}X$ and*

$$[F|_{k,\chi}X] = [F]|_{k,\chi}X.$$

It follows from the theorem that if the genus operator is defined for an eigenform F which is defined by the corresponding eigenvalues up to a constant factor then the following alternative holds:

$$[F] = F \qquad \text{or} \qquad [F] = 0.$$

If would be interesting to verify (or to disprove) it by numerical computations of Fourier coefficients of modular forms.

References

[1] Андрианов А. Н., Малолеткин Г. Н., Поведение тета-рядов рода *n* при модулярных подстановках, *Известия АН СССР (сер. матем.)* **39, 2** (1975), pp. 243–258.

[2] Евдокимов С. А., Базис из собственных функций операторов Гекке в теории модулярных форм рода *n*, *Матем. сборник* **115, 3** (1981), pp. 337–362.

[3] Satake I., Theory of spherical functions on reductive algebraic groups over $\mathfrak{p}$-adic fields, *Publ. Math. I. H. E. S.* **18** (1963), pp. 5–69.

[4] Андрианов А. Н., Эйлеровы разложения тета-преобразований зигелевых модулярных форм рода *n*, *Матем. сборник* **105, 3** (1978), pp. 291–341.

[5] Андрианов А. Н., Мультипликативная арифметика зигелевых модулярных форм, *Успехи матем. наук* **34, 1** (1979), pp. 67–135.

[6] Андрианов А. Н., Эйлеровы произведения отвечающие модулярным формам Зигеля рода 2, *Успехи матем. наук* **29, 3** (1974), pp. 43–110.

[7] Евдокимов С. А., Эйлеровы произведения для конгруэнц-подгруппы зигелевой группы рода 2, *Матем. сборник* **99** (1976), pp. 483-513.

[8] Андрианов А. Н., Калинин В. Л., Об аналитических свойствах стандартных дзета-функций зигелевых модулярных форм, *Матем. сборник* **106, 3** (1978), pp. 323–339.

[9] Freitag E., Die Invarianz gewisser von Thetareihen erzeugter Vektorraume unter Heckeoperatoren, *Math. Z.* **156** (1977), pp. 141–155.

[10] Freitag E., *Eine Bemerkung zu Andrianovs expliziten Formeln für die Wirkung der Heckeoperatoren auf Thetareihen*, E. B. Christoffel, Ed. by P. L. Butzer and F. Fehér, Aachen, Birkhauser Verlag, Basel, 1981.

[11] Andrianov A. N., Action of Hecke operator $T(p)$ on theta series, *Math. Ann.* **247** (1980), pp. 245–254.

[12] Андрианов А. Н., Операторы Гекке и представления бинарных квадратичных форм, *Труды Матем. инст. АН СССР* **165** (1984), pp. 4–15.

[13] Андрианов А. Н., Вырождения колец операторов Гекке на пространствах тета-рядов, *Труды Матем. ин-та АН СССР* **157** (1981), pp. 3–18.

[14] Андрианов А. Н., Двойственность в теореме Зигеля о представлениях родом квадратичных форм и оператор усреднения, *Матем. сборник* **122**(1983), pp. 3–11.

MATHEMATICAL INSTITUTE
FONTANKA 27
LENINGRAD D-11
USSR

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

JEAN-MARC FONTAINE

Représentations p-adiques

On sait que, lorsque l'on étudie les représentations l-adiques du groupe de Galois d'un corps de nombres, les places premières à l jouent un rôle très différent de celles qui divisent l. Tout l'art est dans la manière d'utiliser simultanément les informations fournies par les unes et les autres. Ici, bien au contraire, nous nous limitons à passer en revue quelques résultats sur ce que l'on peut dire lorsque l'on ne considère qu'une seule place $\mathfrak{p}$, supposée en outre diviser l.

Aussi le contexte est le suivant: on fixe un corps K de caractéristique 0, complet pour une valuation discrète, à corps résiduel parfait k de caractéristique $p > 0$; on choisit une clôture algébrique $\overline{K}$ de K et on pose $\mathfrak{G} = \mathrm{Gal}(\overline{K}/K)$. Une *représentation p-adique* est un $\mathbf{Q}_p$-espace vectoriel de dimension finie muni d'une action linéaire et continue de $\mathfrak{G}$.

Comme le groupe $\mathfrak{G}$ n'est pas très "explicite", l'un des objectifs est d'associer à une représentation p-adique des invariants plus tangibles. Le plus naturel est le couple (V, G_V) formé du $\mathbf{Q}_p$-espace vectoriel sous-jacent V et du sous-groupe (fermé) G_V de $\mathbf{GL}(V)$ qui est l'image de $\mathfrak{G}$. On obtient un objet plus maniable en remplaçant G_V

— soit par son algèbre de Lie, $\mathrm{Lie}\, G_V$ (comme tout sous-groupe fermé de $\mathbf{GL}(V)$, G_V est un groupe de Lie p-adique),

— soit par sa clôture de Zariski $G_{V,\mathrm{alg}}$, i.e. le plus petit sous-groupe algébrique de $\mathbf{GL}(V)$ qui contient G_V (d'après un résultat classique de Chevalley, $\mathrm{Lie}\, G_{V,\mathrm{alg}}$ est la plus petite sous-algèbre de Lie algébrique de $\mathfrak{gl}(V)$ qui contient $\mathrm{Lie}\, G_V$).

Nous allons associer à certains types de représentations d'autres invariants et, dans certains cas, voir quels renseignements la connaissance de ces invariants fournit sur $\mathrm{Lie}\, G_V$ ou sur $G_{V,\mathrm{alg}}$.

Dans ce qui suit, mis à part quelques résultats sur les variétés abéliennes et les groupes p-divisibles, on a laissé de côté toutes les questions où intervient la géométrie algébrique (problèmes de comparaison entre différentes cohomologies p-adiques, voir [2], [3], [14], [7] App., [9], bien que ce soit celle-ci qui "motive" beaucoup des définitions données.

Remarque. Beaucoup des catégories abéliennes que nous allons rencontrer sont des $\otimes$-catégories, i.e. sont munies d'un produit tensoriel, d'un Hom interne et d'un objet-unité satisfaisant à des propriétés convenables (cf. [18]); pour alléger l'exposé, nous avons renoncé à donner les définitions, presque toujours évidentes, de ces structures; une sous-$\otimes$-catégorie d'une $\otimes$-catégorie abélienne est une sous-catégorie pleine stable par sous-objet, quotient, somme directe, produit tensoriel, hom interne, contenant l'objet-unité; un $\otimes$-foncteur (resp. une $\otimes$-équivalence de catégories) est ce que l'on pense (voir [18] ou [4] pour des définitions précises).

1. Représentations non ramifiées

On sait depuis longtemps que leur étude se ramène à un problème d'algèbre linéaire. Commençons par fixer quelques notations:

— si A est un anneau commutatif contenant $\mathbf{F}_p$, $W(A)$ est l'anneau des vecteurs de Witt à coefficients dans A et F: $W(A) \to W(A)$ le Frobenius (cf., par exemple, [22], chap. II, § 6);

— on pose $W = W(k)$, $K_0 = \operatorname{Frac} W$ et on note σ le Frobenius agissant sur k, W et K_0;

— on note $W[F]$ (resp. $K_0[F]$) l'anneau (non commutatif si $k \neq \mathbf{F}_p$) engendré par W (resp. K_0) et un élément F, soumis aux relations $F\lambda = \sigma(\lambda)F$, pour tout $\lambda \in W$ (resp. K_0); si A est une k-algèbre, $W(A)$ est, de façon naturelle, une W-algèbre et un $W[F]$-module; si A est parfait, c'est même un $W[F, F^{-1}]$-module.

Par définition, une *représentation non ramifiée* est une représentation sur laquelle le *sous-groupe d'inertie* $\mathfrak{G}_0$ de $\mathfrak{G}$ opère trivialement. Si $\bar{k}$ désigne le corps résiduel de $\bar{K}$, $\bar{k}$ est une clôture algébrique de k et $\mathfrak{G}/\mathfrak{G}_0$ s'identifie à $\tilde{\mathfrak{G}} := \operatorname{Gal}(\bar{k}/k)$.

Le groupe $\tilde{\mathfrak{G}}$ opère sur $P := \operatorname{Frac} W(\bar{k})$ de façon compatible avec sa structure de K_0-algèbre et de $K_0[F, F^{-1}]$-module. Si V est une représentation p-adique non ramifiée, $D_P(V) := (P \otimes_{\mathbf{Q}_p} V)^{\tilde{\mathfrak{G}}}$ est un *module de Dieudonné de pente* 0 *sur* K_0, i.e. un $K_0[F]$-module, de dimension finie comme K_0-espace vectoriel, contenant un W-réseau sur lequel l'action de F est bijective. Le foncteur D_P définit *une $\otimes$-équivalence entre la catégorie des représentations p-adiques non ramifiées et celle des modules de Dieudonné de pente* 0. Un quasi-inverse est le foncteur V_P qui, à un tel module D, associe le sous-$\mathbf{Q}_p[\tilde{\mathfrak{G}}]$-module de $P \otimes_{K_0} D$ formé des éléments fixés par F (ces résultats se démontrent sans difficulté, par exemple en utilisant [23], III, A.1).

2. Une "description" des représentations p-adiques (voir [12], [13], [27], [28] pour des démonstrations et d'autres résultats)

Soit χ: $\mathfrak{G} \to \mathbf{Z}_p^*$ le *caractère cyclotomique*, défini par

$$g\varepsilon = \varepsilon^{\chi(g)}, \quad \text{pour} \quad g \in \mathfrak{G}, \ \varepsilon \in \mu_{p^\infty}(\bar{K});$$

soient $\mathfrak{H}$ le noyau de $\mathfrak{G}$, $\Gamma = \mathfrak{G}/\mathfrak{H}$ et $L = \bar{K}^{\mathfrak{H}}$. Il se trouve que $\mathfrak{H}$ s'identifie au groupe de Galois d'un corps parfait de caractéristique p.

Pour toute extension E de K contenue dans $\bar{K}$, notons R_E l'ensemble des suites $(x_n)_{n\in\mathbf{N}}$ formées d'éléments du complété $\hat{E}$ de E, vérifiant $x_{n+1}^p = x_n$, pour tout n. Si l'on pose

$$(x_n)_{n\in\mathbf{N}} + (y_n)_{n\in\mathbf{N}} = \big(\lim_{m\to\infty}(x_{n+m}+y_{n+m})^{p^m}\big)_{n\in\mathbf{N}},$$
$$(x_n)_{n\in\mathbf{N}} \cdot (y_n)_{n\in\mathbf{N}} = (x_n y_n)_{n\in\mathbf{N}},$$

R_E devient un *corps parfait de caractéristique p*. C'est même un *corps valué complet* (on obtient une valuation de R_E en choisissant une valuation v de $\hat{E}$ et en posant $v\big((x_n)_{n\in\mathbf{N}}\big) = v(x_0)$) dont le corps résiduel s'identifie au corps résiduel k_E de E (bien sûr $R_E = k_E$ si E est une extension finie d'une extension non ramifiée).

Le corps R_L a une structure très simple: c'est *le complété de la clôture radicielle* d'un corps local bien défini, le *corps des normes* $X_K(L)$ de l'extension L/K ([28]), qui est donc un corps de séries formelles en une variable π à coefficients dans k_L. Le groupe abélien Γ opère continûment sur $X_K(L)$ et R_L et cette action est facile à décrire (lorsque p est une uniformisante de K, i.e. lorsque $K = K_0$, on peut choisir π pour que $g(1+\pi) = (1+\pi)^{\chi(g)}$, pour tout $g \in \mathfrak{G}$).

Si maintenant E est une extension finie galoisienne de L, *R_E est une extension finie galoisienne de R_L et* $\mathrm{Gal}(R_E/R_L)$ *s'identifie à* $\mathrm{Gal}(E/L)$; en outre la réunion $R_{\bar{K}}^0$ des R_E, pour E parcourant les extensions finies de L contenues dans $\bar{K}$ est une clôture algébrique de R_E.

Si V est une représentation p-adique, et si $P_R = \mathrm{Frac}\, W(R_{\bar{K}}^0)$, on voit que $D_{P_R}(V) := (P_R \otimes_{\mathbf{Q}_p} V)^{\mathfrak{H}}$ est un "*Γ-module de Dieudonné de pente* 0", i.e. un module de Dieudonné de pente 0 sur $\mathrm{Frac}\, W(R_L)$, munie d'une action continue de Γ, semi-linéaire par rapport à son action naturelle sur $\mathrm{Frac}\, W(R_L)$. On obtient ainsi *une $\otimes$-équivalence entre la catégorie de toutes les représentations p-adiques et celle des Γ-modules de Dieudonné de pente* 0.

3. Décomposition de Hodge–Tate et théorie de Sen ([26], [24], [21])

Si V est une représentation p-adique finie (i.e. telle que $\mathfrak{G}$ opère à travers un groupe fini), la trivialité de $H^1(\mathfrak{G}, \mathbf{GL}_n(\bar{K}))$, pour tout n, implique que $(\bar{K} \otimes_{\mathbf{Q}_p} V)^{\mathfrak{G}}$ est un K-espace vectoriel de dimension finie égale à celle de V sur $\mathbf{Q}_p$. Plus généralement tout $\bar{K}$-espace vectoriel V, de dimension finie n, muni d'une action de G, semi-linéaire par rapport à son action sur $\bar{K}$, continue pour la topologie de Krull de G et *la topologie discrète* de V, est isomorphe à $\bar{K}^n$.

L'idée fondamentale, qui est à la base de toute la suite de cet exposé, et que Tate a été le premier à utiliser dans ce contexte [26], est que, pour les représentations p-adiques quelconques, il faut remplacer $\bar{K}$ par son complété $C = \hat{\bar{K}}$ (et donc $H^1(\mathfrak{G}, \mathbf{GL}_n(\bar{K}))$ par $H^1_{\mathrm{cont}}(\mathfrak{G}, \mathbf{GL}_n(C))$, la topologie de $\mathbf{GL}_n(C)$ étant la topologie p-adique). Autrement dit, dans un premier temps, au lieu d'étudier les représentations p-adiques, on va s'intéresser aux *C-représentations* de $\mathfrak{G}$, i.e. aux C-espaces vectoriels de dimension finie, munis d'une action semi-linéaire et continue de $\mathfrak{G}$.

Le premier résultat, dû à Tate, est le suivant: si $\mathscr{V}$ est une C-représentation de $\mathfrak{G}$, et si, pour tout $i \in \mathbf{Z}$, on note $\mathscr{V}^i$ le sous-K-espace vectoriel de $\mathscr{V}$ formé des v vérifiant

$$gv = \chi^i(g)\cdot v, \qquad \text{pour tout } g \in \mathfrak{G},$$

alors ([24], prop. 4), *l'application évidente*

$$\bigoplus_{i\in\mathbf{Z}} C \otimes_K \mathscr{V}^i \to \mathscr{V}$$

est injective. Si elle est surjective, on a alors une décomposition canonique de $\mathscr{V}$, *la décomposition de Hodge–Tate*

$$\mathscr{V} = \bigoplus_{i\in\mathbf{Z}} C \otimes_K \mathscr{V}^i,$$

et on dit que $\mathscr{V}$ est *du type Hodge–Tate*. L'un des intérêts de cette notion est que l'on conjecture (Tate) que la cohomologie étale p-adique fournit, par extension des scalaires à C, des exemples de C-représentations du type Hodge–Tate.

Une classification pratiquement complète des C-représentations a été obtenue par Sen ([21]): La théorie de Sen *associe, à toute C-représentation $\mathscr{V}$ de dimension n, un L-espace vectoriel $\mathscr{W}$ de dimension n muni d'un endomorphisme φ, dont les facteurs invariants sont à coefficients dans K. La connaissance de φ détermine $\mathscr{V}$ à isomorphisme près; si k est algébriquement clos, tout couple $(\mathscr{W}, \varphi)$ provient d'une C-représentation.* En termes de cocycles continus, cela revient à construire une application injective (bijective si k est algébriquement clos) de $H^1_{\text{cont}}(\mathfrak{G}, \mathbf{GL}_n(C))$ dans l'ensemble $\tilde{\mathbf{M}}_n(K)$ des classes de similitude des matrices (n, n) à coefficients dans K. La construction de $(\mathscr{W}, \varphi)$, qui se fait en trois étapes, est instructive:

1° *étape*: L'ensemble $H^1_{\text{cont}}(\mathfrak{H}, \mathbf{GL}_n(C))$ a un seul élément. Comme en outre $C^{\mathfrak{H}} = \hat{L}$ (le complété de $L = \bar{K}^{\mathfrak{H}}$), pour toute C-représentation $\mathscr{V}$ de $\mathfrak{G}$, l'application naturelle de $C \otimes_{\hat{L}} V^{\mathfrak{H}}$ dans V est un isomorphisme; on est donc ramené à l'étude "des $\hat{L}$-représentations" de Γ.

2° *étape*: Il suffit en fait d'étudier "les L-représentations" de Γ, i.e. les L-espaces vectoriels de dimension finie munis d'une action semi-linéaire et continue de Γ. On peut en effet "décompléter" la $\hat{L}$-représentation $\mathscr{V}^{\mathfrak{H}}$: si $\mathscr{W}$ désigne le sous-L-espace vectoriel de $\mathscr{V}^{\mathfrak{H}}$, réunion des sous-$L$-espaces vectoriels de dimension finie stables par Γ, il est lui-même de dimension finie et l'application naturelle de $\hat{L} \otimes_L \mathscr{W}$ dans $\mathscr{V}^{\mathfrak{H}}$ est un isomorphisme.

3° *étape*: Choisissons une base $e_1, e_2, \ldots, e_n$ de $\mathscr{W}$ sur L. Il existe une extension finie K' de K contenue dans L telle que le sous-K'-espace vectoriel $\mathscr{W}'$ de $\mathscr{W}$ engendré par les e_j est stable par Γ. L'action de $\mathrm{Gal}(L/K')$ sur $\mathscr{W}'$ est alors *linéaire* et l'on en déduit facilement l'existence d'un unique L-endomorphisme φ de $\mathscr{W}$ tel que

$$\gamma w = \exp(\varphi \cdot \log \chi(\gamma))(w), \tag{$*$}$$

pour tout $w \in \mathscr{W}'$ et tout γ appartenant à un sous-groupe ouvert suffisamment petit de Γ. Il est clair que φ ne dépend pas de la base choisie, parce que la formule $(*)$ reste vraie pour tout $w \in \mathscr{W}$ et tout γ appartenant à un sous-groupe ouvert convenable (dépendant de w) de Γ. On montre alors que l'on peut choisir une base sur laquelle la matrice de φ est à coefficients dans K.

Remarques. 1. Une C-représentation est du type Hodge–Tate si et seulement si φ est semi-simple, et à valeurs propres dans $\mathbf{Z}$.

2. L'endomorphisme φ ne dépend que de la restriction de l'action de $\mathfrak{G}$ à un sous-groupe ouvert du groupe d'inertie: Le foncteur qui à $\mathscr{V}$ associe le couple $(\mathscr{W}, \varphi)$ est un $\otimes$-foncteur exact et fidèle de la catégorie des C-représentations dans celle des L-espaces vectoriels de dimension finie, munis d'un endomorphisme; mais, bien que $(\mathscr{W}, \varphi)$ détermine $\mathscr{V}$ à isomorphisme près, il n'est pas pleinement fidèle. Si k est algébriquement clos, il l'est "presque": il induit une $\otimes$-équivalence entre, d'une part, la catégorie dont les objets sont les C-représentations et les morphismes

les applications C-linéaires qui commutent à l'action d'un sous-groupe ouvert suffisamment petit de $\mathfrak{G}$ et, d'autre part, la catégorie des couples $(\mathscr{W}, \varphi)$.

4. Représentations linéaires de Lie$G$ [(24], [20], [21])

Si V est une représentation p-adique, on peut appliquer la théorie de Sen à la C-représentation $V_C = C \otimes_{\mathbf{Q}_p} V$ et V_C est muni d'un C-endomorphisme φ_{V_C} (c'est l'extension C-linéaire à V_C de l'endomorphisme L-linéaire φ de $\mathscr{W}$; on prendra garde que, bien que $V_C = C \otimes_{\mathbf{Q}_p} V = C \otimes_L \mathscr{W}$, en général $\mathscr{W}$ est distinct de $L \otimes_{\mathbf{Q}_p} V$). D'où la question: que peut-on dire de l'action de $\mathfrak{G}$ sur V si l'on connaît seulement φ_{V_C}? La réponse, conjecturée par Serre dans le cas où V_C est du type Hodge–Tate, a été obtenue par Sen: la connaissance de φ_{V_C} est équivalente à celle de l'action de l'algèbre de Lie du groupe d'inertie. En particulier, *si l'on suppose k algébriquement clos,* LieG_V *est la plus petite sous-algèbre de Lie* (ou le plus petit sous-$\mathbf{Q}_p$-espace vectoriel, il se trouve que cela revient au même) *de* $\mathfrak{gl}(V)$ *défini(e) sur* $\mathbf{Q}_p$, *dont l'extension des scalaires à C contient* φ_{V_C}.

Remarques. 1. Supposons toujours k algébriquement clos et soit $\mathfrak{g} = \mathrm{Lie}\,\mathfrak{G} := \varprojlim \mathrm{Lie}(\mathfrak{G}/\mathfrak{J})$, pour $\mathfrak{J}$ parcourant les sous-groupes fermés invariants de $\mathfrak{G}$ tels que $\mathfrak{G}/\mathfrak{J}$ est un groupe de Lie p-adique de dimension finie. Soit $\mathfrak{g}_C = C \hat{\otimes}_{\mathbf{Q}_p} \mathfrak{g} := \varprojlim C \otimes_{\mathbf{Q}_p} \mathrm{Lie}(\mathfrak{G}/\mathfrak{J})$. Il existe un unique élément $\varphi_{\mathrm{Sen}} \in \mathfrak{g}_C$ tel que, pour toute représentation p-adique V, φ_{V_C} soit l'endomorphisme qui donne l'action de φ_{Sen} sur V_C. Le résultat ci-dessus signifie que $\mathfrak{g}$ est la plus petite sous-algèbre de Lie d'elle-même, fermée, définie sur $\mathbf{Q}_p$, et dont l'extension des scalaires à C contient φ_{Sen}.

2. On sait peu de choses sur ce que peuvent être les facteurs invariants d'un φ_{V_C} provenant d'une représentation p-adique V. Il ne paraît pas impossible de pouvoir les calculer en termes du Γ-module de Dieudonné de pente 0 associé à V.

5. Représentations de Hodge–Tate, de de Rham et cristallines ([6], [7], [9]–[11], [16], [29])

Posons $\mathbf{Q}_p(1) = \mathbf{Q}_p \otimes_{\mathbf{Z}_p} \varprojlim \mu_{p^\infty}(\bar{K})$; pour tout $i \in \mathbf{N}$, soit $\mathbf{Q}_p(i)$ la i-ième puissance tensorielle de $\mathbf{Q}_p(i)$ et $\mathbf{Q}_p(-i)$ le dual de $\mathbf{Q}_p(i)$. Pour $i \in \mathbf{Z}$, $\mathbf{Q}_p(i)$ est un $\mathbf{Q}_p$-espace vectoriel de dimension 1 sur lequel $\mathfrak{G}$ opère à travers χ^i.

Notons B_{HT} la C-algèbre graduée $\bigoplus_{i\in\mathbf{Z}} C\otimes_{\mathbf{Q}_p}\mathbf{Q}_p(i)$ (si t est un élément non nul de $\mathbf{Q}_p(1)$, tout élément de B_{HT} s'écrit, de manière unique, sous la forme $\sum_{i\in\mathbf{Z}} c_i t^i$, avec les $c_i\in C$, presque tous nuls).

Pour toute représentation p-adique V, soit $D_{\mathrm{HT}}(V)=(B_{\mathrm{HT}}\otimes_{\mathbf{Q}_p}V)^{\mathfrak{G}}$. D'après les résultats de Tate rappelés au n° précédent, c'est un K-espace vectoriel gradué de dimension finie inférieure ou égale à la dimension de V sur $\mathbf{Q}_p$; on dit que V est *de Hodge–Tate* si ces deux dimensions sont égales (ce qui revient à dire que V_C est du type Hodge–Tate). On montre facilement que les représentations de Hodge–Tate forment une sous-$\otimes$-catégorie de celle des représentations p-adiques et que, par restriction, D_{HT} induit *un $\otimes$-foncteur exact et fidèle de cette catégorie sur celle des K-espaces vectoriels gradués de dimension finie.*

Soit R l'anneau des entiers de $R_{\overline{K}}$ (cf. n° 2, $R_{\overline{K}}$ est le complété de $R^0_{\overline{K}}$) et soit $W_K(R)=K\otimes_W W(R)$. Si $a=(a_0,a_1,\ldots,a_m,\ldots)\in W(R)$, chaque a_m est une suite d'éléments $(a_{m,n})_{n\in\mathbf{N}}$ d'éléments de l'anneau des entiers $\mathcal{O}_C$ de C; l'application $\theta^0\colon W(R)\to\mathcal{O}_C$, qui à a associe $\sum_{m=0}^{\infty}p^m a_{m,m}$, est un homomorphisme surjectif de W-algèbres, qui induit un épimorphisme, encore noté θ^0, de la K-algèbre $W_K(R)$ sur C; son noyau est donc un idéal maximal $\mathfrak{m}$ de $W_K(R)$ et on note B_{DR} le corps des fractions du séparé complété de $W_K(R)$ pour la topologie $\mathfrak{m}$-adique. C'est un corps complet pour une valuation discrète, dont le corps résiduel s'identifie à C; on montre que $\mathbf{Q}_p(1)$ se plonge canoniquement dans le groupe additif de B_{DR} et que tout élément non nul de $\mathbf{Q}_p(1)$ est une uniformisante de B_{DR}; il en résulte que l'anneau gradué associé à B_{DR} (pour la filtration définie par les puissances de l'idéal maximal de l'anneau des entiers de B_{DR}) s'identifie à B_{HT}.

Pour toute représentation p-adique V, soit $D_{\mathrm{DR}}(V)=(B_{\mathrm{DR}}\otimes_{\mathbf{Q}_p}V)^{\mathfrak{G}}$. Comme $\mathrm{gr}^{\cdot}B_{\mathrm{DR}}=B_{\mathrm{HT}}$, c'est un K-espace vectoriel filtré de dimension finie inférieure ou égale à la dimension de V sur $\mathbf{Q}_p$; on dit que V est *de de Rham* si ces dimensions sont égales; s'il en est ainsi, V est aussi de Hodge–Tate et $\mathrm{gr}^{\cdot}D_{\mathrm{DR}}(V)$ s'identifie à $D_{\mathrm{HT}}(V)$.

Ici encore les représentations de de Rham forment une sous-$\otimes$-catégorie des représentations de Hodge–Tate et, par restriction, D_{DR} induit *un $\otimes$-foncteur exact et fidèle de la catégorie des représentations de de Rham sur celle des K-espaces vectoriels filtrés de dimension finie* (cette dernière n'est pas abélienne, mais on a toutefois des notions de suite exacte courte, produit tensoriel, ...).

Remarque. S'il est clair qu'il existe un homomorphisme de la K-algèbre C dans l'anneau des entiers de B_{DR} qui composé avec la projection de cet anneau sur B_{DR} induise l'identité, je ne vois pas de raison pour que l'on puisse choisir un tel homomorphisme compatible avec l'action de $\mathfrak{G}$. On peut donc penser qu'il existe des représentations de Hodge–Tate qui ne sont pas de de Rham; les extensions non triviales de $\mathbf{Q}_p(1)$ par $\mathbf{Q}_p$ semblent de bons candidats, mais je ne sais pas le montrer (les extensions de $\mathbf{Q}_p$ par $\mathbf{Q}_p(1)$ sont en revanche non seulement de Hodge–Tate mais aussi de de Rham).

Soit $W^{DP}(R)$ le séparé complété pour la topologie p-adique de l'enveloppe à puissances divisées de $W(R)$ relativement au noyau de θ^0 (choisissons un $x = (x_n)_{n\in\mathbf{N}} \in R$ tel que $x_0 = p$ et posons $[x] = (x, 0, 0, \ldots)$, $\xi = [x]-p$, $W_{K_0}(R) = K_0 \otimes_W W(R)$, $\gamma_n(\xi) = \frac{1}{n!}\otimes \xi^n$, et $\gamma_n([x]) = \frac{1}{n!}\otimes \otimes [x^n]$, pour $n\in\mathbf{N}$; alors $W(R)\left[(\gamma_n(\xi))_{n\in\mathbf{N}}\right] = W(R)\left[(\gamma_n([x]))_{n\in\mathbf{N}}\right]$ et $W^{DP}(R)$ est le séparé complété de cette algèbre pour la topologie p-adique).

L'homomorphisme évident de $W_{K_0}(R)$ dans $W_K(R)$ se prolonge de manière naturelle en une inclusion de $W_{K_0}^{DP}(R) := K_0\otimes_W W^{DP}(R)$ dans B_{DR} que nous utilisons pour identifier $W_{K_0}^{DP}(R)$ à un sous-anneau de B_{DR}. On note B_{cris} la sous-$W_{K_0}^{DP}(R)$-algèbre de B_{DR} engendrée par t^{-1}, où t est un élément arbitraire non nul de $\mathbf{Q}_p(1)$.

Il est clair que B_{cris} est stable par $\mathfrak{G}$. On montre en outre

(i) que l'action de F sur $W(R)$ s'étend naturellement à B_{cris},

(ii) que l'homomorphisme évident de $K\otimes_{K_0} B_{\mathrm{cris}}$ dans B_{DR} est injectif.

Pour toute représentation p-adique V, soit

$$D_{\mathrm{cris}}(V) = (B_{\mathrm{cris}}\otimes_{\mathbf{Q}_p} V)^{\mathfrak{G}}.$$

L'injectivité de $K\otimes_{K_0} B_{\mathrm{cris}}$ dans B_{DR} induit une injection de $K\otimes_{K_0} D_{\mathrm{cris}}(V)$ dans $D_{\mathrm{DR}}(V)$ et $D_{\mathrm{cris}}(V)$ est donc un K_0-espace vectoriel de dimension finie inférieure ou égale à celle de V sur $\mathbf{Q}_p$; on dit que V est *cristalline* si ces dimensions sont égales.

Si V est cristalline, V est aussi de de Rham et $D_{\mathrm{DR}}(V)$ s'identifie à $K\otimes_{K_0} D_{\mathrm{cris}}(V)$. S'il en est ainsi, $D_{\mathrm{cris}}(V)$ est un *module de Dieudonné filtré faiblement admissible* au sens de [13]; autrement dit, c'est un K_0-espace vectoriel D de dimension finie muni

(i) d'une action σ-semi-linéaire de F (induite par l'action de F sur B_{cris}),

(ii) d'une filtration de $D_K = K \otimes_{K_0} D$ (la filtration naturelle de $D_{\mathrm{DR}}(V)$), vérifiant certaines conditions (voir [6], § 4 et [16], § 1).

La catégorie $\underline{\mathrm{MF}}_K^f$ des modules de Dieudonné filtrés faiblement admissibles est abélienne. On dit qu'un module de Dieudonné filtré est *admissible* s'il est isomorphe à un $D_{\mathrm{cris}}(V)$, pour une représentation cristalline V convenable. La catégorie $\underline{\mathrm{MF}}_K^a$ des modules de Dieudonné filtrés admissibles est une sous-catégorie pleine de $\underline{\mathrm{MF}}_K^f$ stable par sous-objet et quotient.

On montre que *la restriction de D_{cris} à la catégorie des représentations cristallines est pleinement fidèle et induit une $\otimes$-équivalence de cette catégorie sur $\underline{\mathrm{MF}}_K^a$.*

Le défaut de cette théorie est que l'on ne connaît pas de description explicite de $\underline{\mathrm{MF}}_K^a$; mais, en fait, on conjecture que faiblement admissible équivaut à admissible et on sait le démontrer dans de nombreux cas particuliers ([15], [16], [11], [10]).

Remarques. 1. L'anneau B_{cris} construit ici diffère légèrement de celui construit dans [7] qu'il contient. Mais cela ne change pas le foncteur D_{cris}. Si $\mathcal{O}_{\overline{K}}$ désigne l'anneau des entiers de $\overline{K}$, $W^{DP}(R)$ s'identifie (cf. [10]) à

$$H^0_{\mathrm{cris}}(\mathcal{O}_{\overline{K}}) := \varprojlim H^0\Big(\big(\mathrm{Spec}(\mathcal{O}_{\overline{K}}/p^n\mathcal{O}_{\overline{K}})/W_n(k)\big)_{\mathrm{cris}},\ \text{faisc. struct.}\Big).$$

2. Si H est un *groupe de Barsotti–Tate* sur l'anneau des entiers de K, $V_p(H) = \mathbf{Q}_p \otimes_{\mathbf{Z}_p} \varprojlim H_{p^n}(\mathcal{O}_{\overline{K}})$ est une représentation cristalline de poids $\subset \{0,1\}$ (i.e. $V_C^i \neq 0 \Rightarrow i \in \{0,1\}$); inversement, si $e = [K:K_0] \leqslant p-1$, toute représentation cristalline de poids $\subset \{0,1\}$ provient d'un groupe de Barsotti–Tate ([17], [5], [16]); on conjecture que cela reste vrai si $e \geqslant p$ (on le sait dans des cas particuliers, voir [15], [10]).

3. Si A est une *variété abélienne* sur K, $V_p(A) = \mathbf{Q}_p \otimes_{\mathbf{Z}_p} \varprojlim A_{p^n}(\overline{K})$ est de Hodge–Tate de poids $\subset \{0,1\}$ ([26], [1], [8]) et même de de Rham. Elle est cristalline si (et probablement seulement si, c'est en tout cas un théorème si $e \leqslant p-1$) elle a bonne réduction.

4. Il serait agréable de savoir reconnaître si une représentation p-adique est de Hodge–Tate, de de Rham ou cristalline en termes de son Γ-module de Dieudonné de pente 0. Il serait agréable d'avoir aussi une description terre à terre du foncteur qui associe à tout module de Dieudonné filtré admissible le Γ-module de Dieudonné de pente 0 de la représentation cristalline correspondante. Les anneaux $\mathrm{Frac}\, W(R^0_{\overline{K}})$ et B_{cris} sont suffisamment "voisins" pour que cela paraisse possible.

6. Le groupe $G_{V,\mathrm{alg}}$ pour les représentations de Hodge–Tate ([23]–[25], [20], [21], [28], [29])

On suppose maintenant k algébriquement clos.

Soit V une représentation p-adique. Si φ_{V_C} est semi-simple et si le $\mathbf{Q}$-espace vectoriel engendré par ses valeurs propres est de dimension < 1, le sous-C-espace vectoriel de $\mathfrak{gl}(V_C)$ engendré par φ_{V_C} est une sous-algèbre de Lie algébrique et $\mathrm{Lie}\, G_{V,\mathrm{alg}} = \mathrm{Lie} G_V$ (en particulier, G_V est ouvert dans $G_{V,\mathrm{alg}}(\mathbf{Q}_p)$).

C'est le cas lorsque V est de Hodge–Tate. Soit alors $h_V\colon \mathbf{G}_m \to \mathbf{GL}(V_C)$ l'homomorphisme défini par $h_V(\lambda)\cdot x = \lambda^i x$, si $x \in V_C^i$; *l'image de h_V est contenue dans $G_{V,\mathrm{alg}} \otimes C$* et *la classe de conjugaison $\mathscr{C}_V$ de h_V dans $G_{V,\mathrm{alg}}$ est définie sur K*. On a donc un triplet $(G, \mathscr{C}, U)$ (avec $G = G_{V,\mathrm{alg}}$, $\mathscr{C} = \mathscr{C}_V$, $U = V$) formé d'un groupe algébrique G défini sur $\mathbf{Q}_p$, d'une classe de conjugaison $\mathscr{C}$ de sous-groupes à un paramètre de G, définie sur K et d'une représentation linéaire U de dimension finie de G. Si on suppose que V est à poids $\subset [j, j']$ (i.e. que $V_C^i \neq 0$ implique $j \leqslant i \leqslant j'$), le triplet $(G, \mathscr{C}, U)$ vérifie

$$\underline{MT}_{[j,j']}\begin{cases}\text{(i) } \textit{la composante neutre } G^0 \textit{ de } G \textit{ est le plus petit sous-groupe} \\ \quad \textit{algébrique de } G\textit{, défini sur } \mathbf{Q}_p\textit{, contenant } \mathscr{C}, \\ \text{(ii) } U \textit{ est une représentation fidèle de } G \textit{ et les poids de l'action de} \\ \quad \mathscr{C} \textit{ sur } U \textit{ sont compris entre } j \textit{ et } j'.\end{cases}$$

On peut tenter de classifier de tels triplets, au moins lorsque l'on suppose G réductif. Cela a été fait par Serre ([25]) pour $\underline{MT}_{[0,1]}$: il montre en particulier que *les facteurs simples de G^0 sont*, dans ce cas, *de type classique* ($\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n, \mathbf{D}_n$) et que *leurs poids dans U sont des poids minuscules.*

Soit $(G, \mathscr{C}, U)$ un triplet vérifiant $\underline{MT}_{[j,j']}$. Existe-t-il une représentation V de Hodge–Tate telle que $(G_{V,\mathrm{alg}}, \mathscr{C}_V, V) \simeq (G, \mathscr{C}, U)$? On dispose de méthodes pour attaquer ce problème si l'on exige en plus que V soit cristalline. Le groupe G doit alors être *connexe* ([6]), mais ce n'est pas suffisant. Si G est *réductif connexe*, les résultats partiels dont on dispose laissent penser que la réponse pourrait être *oui*.

Supposons $e = 1$, i.e. $K = K_0$ (c'est le seul cas où les résultats sont significatifs). Un travail de Wintenberger ([29]) permet d'associer, à tout module de Dieudonné filtré faiblement admissible D, un triplet $(G_D, \mathscr{C}_D, U_D)$ (vérifiant $\underline{MT}_{[j,j']}$ si $\mathrm{gr}^{-i} D = 0$ implique $j \leqslant i \leqslant j'$). Lorsque D est admissible, si $D \simeq D_{\mathrm{cris}}(V)$, le fait que D_{cris} définit une $\otimes$-équivalence de catégories implique que les triplets $(G_{V,\mathrm{alg}}, \mathscr{C}_V, V)$ et $(G_D, \mathscr{C}_D, U_D)$ sont des "formes intérieures" l'un de l'autre et deviennent isomorphes après une extension finie non ramifiée de $\mathbf{Q}_p$.

Dans [30], Wintenberger montre que, quelque soient j et j', *tout triplet* $(G, \mathscr{C}, U)$ *vérifiant* $\underline{MT}_{[j,j']}$, *avec* G *réductif connexe, est isomorphe à un* $(G_D, \mathscr{C}_D, U_D)$, *pour un* D *faiblement admissible convenable.*

Comme on sait ([11]) que *tout* D *faiblement admissible, tel qu'il existe* j *avec* $\mathrm{Fil}^j D = D$ *et* $\mathrm{Fil}^{j+p} D = 0$, *est admissible, cela implique que, si* $j' - j < p$, *tout* $(G, \mathscr{C}, U)$ *vérifiant* $\underline{MT}_{[j,j']}$, *avec* G *réductif connexe, est isomorphe à une forme intérieure du triplet associé à une représentation cristalline de poids* $\subset [j, j']$ *convenable.*

Remarques. 1. Ces résultats joints à ceux de Serre pour $\underline{MT}_{[0,1]}$ et à la remarque 2 du n° 3 permettent, *lorsque* $e = 1$, de *caractériser, à une forme intérieure près, les triplets* $(G, \mathscr{C}, U)$ *avec* G *réductif, que l'on obtient à partir des représentations p-adiques de la forme* $V_p(A) = \mathbf{Q}_p \otimes_{\mathbf{Z}_p} \varprojlim A_{p^n}(\bar{K})$, *où* A *est un groupe p-divisible ou un schéma abélien sur les entiers de* K (cf. [30]).

2. Cela fournit un moyen pour montrer que n'importe quel groupe réductif connexe peut se réaliser (à torsion par une forme intérieure près) comme $G_{V,\mathrm{alg}}$ d'une représentation cristalline V, du moins si p est assez grand; par exemple $\mathbf{SL}_2$ (si $p \geqslant 3$) et (Serre) $\mathbf{G}_2$ (si $p \geqslant 3$), $\mathbf{E}_8$ (si $p \geqslant 5$); il est intéressant de noter qu'aucun de ces trois groupes n'est le $G_{V,\mathrm{alg}}$ d'une représentation appartenant à la $\otimes$-catégorie engendrée par les modules de Tate des groupes p-divisibles et leurs duaux (cf. [25], th. 7).

Bibliographie

[1] Bogomolov F. A., Sur l'algébricité des représentations l-adiques, *C. R. Acad. Sci. Paris* **290** (1980), pp. 701–703.

[2] Bloch S. and Kato K., *p-adic Etale Cohomology*, à paraître.

[3] Berthelot P. and Ogus A., F-isocrystals and de Rham Cohomology I, *Inv. Math.* **72** (1983), pp. 159–199.

[4] Deligne P. and Milne J., Tannakian Categories. In: *Hodge Cycles, Motives and Shimura Varieties*, Lecture Notes in Math. **900**, Springer, Berlin, 1982.

[5] Fontaine J.-M., Groupes p-divisibles sur les corps locaux, *Astérisque* **47–48**, Soc. math. de France, 1977.

[6] Fontaine J.-M., Modules galoisiens, modules filtrés et anneaux de Barsotti–Tate. In: *Journées de Géométrie Algébrique de Rennes, Astérisque* **65**, Soc. math. de France, 1979.

[7] Fontaine J.-M., Sur certains types de représentations p-adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti–Tate, *Ann. of Math.* **115** (1982), pp. 529–577.

[8] Fontaine J.-M., Formes différentielles et modules de Tate des variétés abéliennes sur les corps locaux, *Inv. Math.* **65** (1982), pp. 379–409.

[9] Fontaine J.-M., Cohomologie de de Rham, cohomologie cristalline et représentations p-adiques. In: *Algebraic Geometry, Proceedings, Tokyo/Kyoto 1982*, Lecture Notes in Math. **1016**, Springer, Berlin, 1983.

[10] Fontaine J.-M., *Représentations p-adiques ordinaires*, en préparation.

[11] Fontaine J.-M. et Laffaille G., Construction de représentations p-adiques, *Ann. Sci. E.N.S.* **15** (1982), pp. 547–608.

[12] Fontaine J.-M. et Wintenberger J.-P., Le "corps des normes" de certaines extensions algébriques de corps locaux, *C.R.A.S.* **288** (1979), pp. 367–370.

[13] Fontaine J.-M. et Wintenberger J.-P., Extensions algébriques et corps des normes des extensions APF des corps locaux, *C.R.A.S.* **288** (1979), pp. 441–444.

[14] Gillet H. et Messing W., article en préparation.

[15] Laffaille G., Construction de groupes p-divisibles: le cas de dimension l. In: *Journées de Géométrie Algébrique de Rennes, Astérisque* **65**, Soc. math. de France, 1979.

[16] Laffaille G., Groupes p-divisibles et modules filtrés: le cas peu ramifié, *Bull. Soc. Math. de France* **108** (1980), pp. 187–206.

[17] Messing W., *The Crystals Associated to Barsotti–Tate Groups: with Applications to Abelian Schemes*, Lecture Notes in Math. **264**, Springer, Berlin, 1972.

[18] Saavedra Rivano N., *Catégories Tannakiennes*, Lecture Notes in Math. **265**, Springer, Berlin, 1972.

[19] Sen S., On Automorphisms of Local Fields, *Ann. of Math.* **90** (1969), pp. 33–46.

[20] Sen S., Lie Algebras of Galois Groups Arising from Hodge–Tate Modules, *Ann. of Math.* **97** (1973), pp. 160–170.

[21] Sen S., Continuous Cohomology and p-adic Galois Representations, *Inv. Math.* **62** (1980), pp. 89–116.

[22] Serre J.-P., *Corps locaux*, 2° éd., Hermann, Paris, 1968.

[23] Serre J.-P., *Abelian l-adic Representations and Elliptic Curves*, Benjamin, New York, 1968.

[24] Serre J.-P., Sur les groupes de Galois attachés aux groupes p-divisibles. In: *Proceedings of a Conference on Local Fields*, Springer, Berlin, 1967.

[25] Serre J.-P., Groupes algébriques associés aux modules de Hodge–Tate. In: *Journées de Géométrie Algébrique de Rennes, Astérisque* **65**, Soc. math. de France, 1979.

[26] Tate J., p-divisible Groups. In: *Proceedings of a Conference on Local Fields*, Springer, Berlin, 1967.

[27] Wintenberger J.-P., Extensions de Lie et groupes d'automorphismes de corps locaux de caractéristique p, *C.R.A.S.* **288** (1979), pp. 477–479.

[28] Wintenberger J.-P., Le corps des normes de certaines extensions infinies de corps locaux; applications, *Ann. Sci. E.N.S.* (4) **16** (1983), pp. 59–89.

[29] Wintenberger J.-P., Un scindage de la filtration de Hodge pour certaines variétés algébriques sur les corps locaux, *Ann. of Math.*, à paraître.

[30] Wintenberger J.-P., *Groupes algébriques associés à certaines représentations p-adiques*, à paraître.

LABORATOIRE DE MATHÉMATIQUES PURES ASSOCIÉ AU C.N.R.S.
INSTITUT FOURIER, UNIVERSITÉ DE GRENOBLE I,
B.P. 74, 38402 SAINT MARTIN D'HÈRES CEDEX, FRANCE

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

D. R. HEATH-BROWN

Finding Primes by Sieve Methods

At the 1978 International Congress Professor Iwaniec mentioned, as an application of the linear sieve, the fact that the difference between consecutive primes p_n satisfies

$$p_{n+1}-p_n \ll p_n \tag{1}$$

for any $\vartheta > \frac{11}{20}$. This result is due to Heath-Brown and Iwaniec [3] and improves an earlier theorem of Iwaniec and Jutila [6] which covered the range $\vartheta > \frac{5}{9}$ only. These results were obtained by combining sieve methods with the existing machinery of the Riemann Zeta-function and zero density estimates. The earlier methods culminated in Huxley's proof [4] that

$$\pi(x+x^{\vartheta})-\pi(x) \sim \frac{x^{\vartheta}}{\log x}, \quad \vartheta > \frac{7}{12}$$

(so that (1) holds for $\vartheta > \frac{7}{12}$, *a fortiori*). The constant $\frac{7}{12}$ is still the limit for such an asymptotic formula; the ranges $\vartheta > \frac{11}{20}$, $\vartheta > \frac{5}{9}$ arise from lower bounds

$$\pi(x+x^{\vartheta})-\pi(x) \geqslant C(\vartheta)\frac{x^{\vartheta}}{\log x} \tag{2}$$

with $0 < C(\vartheta) < 1$.

The purpose of this lecture is to examine recent developments in this circle of ideas. Iwaniec [5] and Pintz [7] independently have extended the range of (1) to $\vartheta > \frac{17}{31} = 0.5483\ldots$, but I shall not go into this; it was already known in 1978 that some improvement on $\frac{11}{20} = 0.55$ must be possible. I shall be concerned with an analysis of the sieve principles involved in such methods and with their applications ot other problems.

Let us first consider the method of Heath-Brown and Iwaniec. One starts with the sets $\mathscr{A} = \{n; x < n \leqslant x+x^{\vartheta}\}$, $\mathscr{A}_k = \{n \in \mathscr{A}; k|n\}$, and considers the sieve function

$$S(\mathscr{A}_k, z) = \#\{n \in \mathscr{A}_k;\, p|n \Rightarrow p \geqslant z\}.$$

Then $S(\mathscr{A}, x^{\frac{1}{2}})$ is essentially the number of primes in $\mathscr{A}$. The Buchstab identity yields

$$S(\mathscr{A}, x^{\frac{1}{2}}) = \#\mathscr{A} - \sum_{q<x^{\frac{1}{2}}} S(\mathscr{A}_q, q), \tag{3}$$

where q runs over primes. Careful use of the linear sieve, employing Iwaniec's bilinear form for the remainder sum, allows one to give upper bounds for $S(\mathscr{A}_q, q)$. This produces an estimate of the form (2), valid for $\frac{11}{20} < \vartheta < 1$. The constant $C(\vartheta)$ is a continuous function of ϑ, but unfortunately satisfies

$$C(\vartheta) = 0 \quad \left(\vartheta \geqslant \frac{7}{12}\right), \qquad C(\vartheta) < 0 \quad \left(\vartheta < \frac{7}{12}\right).$$

To get a positive constant $C(\vartheta)$ one takes a range $Q_1 < q \leqslant Q_2$ in (3) and replaces the sieve upper bound for $S(\mathscr{A}_q, q)$ by an asymptotic formula. This improves the overall result by $C'(\vartheta)x^{\vartheta}(\log x)^{-1}$, where $C'(\vartheta)$ is continuous and positive. It follows that $C(\vartheta)+C'(\vartheta) > 0$ on some interval $\frac{7}{12} - \delta \leqslant \vartheta \leqslant 1$. When $\vartheta < \frac{7}{12}$ there are only certain special ranges $Q_1 < q \leqslant Q_2$ in which asymptotic formulae are at present available.

It is natural to try to develope (3) by repeating the Buchstab identity on those terms $S(\mathscr{A}_q, q)$ which have not been estimated asymptotically. This would produce

$$S(\mathscr{A}_q, q) = \#\mathscr{A}_q - \sum_{r<q} S(\mathscr{A}_{qr}, qr).$$

Again one could deal satisfactorily with certain particular ranges of q and r. The remaining terms, which now have positive sign, may either be bounded below by zero or iterated further. The process is precisely the same as that used to produce the Rosser sieve.

Clearly it is important to identify those sums $\sum_d' S(\mathscr{A}_d, z)$ for which an asymptotic formula can be given. (Here $\sum_d'$ means that d runs over those integers whose prime factors p_i lie in certain specified intervals $P_i < p_i \leqslant 2P_i$.) The method for handling these sums is such that if $\sum_d' S(\mathscr{A}_d, z)$ can be dealt with then so can

$$\sum_d{}' \Big(\#\mathscr{A}_d - \sum_{q<z} S(\mathscr{A}_{qd}, q)\Big).$$

By repeated iterations of Buchstab's formula it therefore suffices to consider $\sum_d' \# \mathscr{A}_d$. Vaughan's identity may be used for the sum over p_i, $P_i < p_i \leqslant 2P_i$, but cannot help when $P_i \leqslant x^{(1-\vartheta)/2}$. One should therefore start from the identity

$$\sum_{n \geqslant z} \frac{\Lambda(n)}{\log n} n^{-s} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \big(\zeta(s)\Pi(s)-1\big)^k, \tag{4}$$
$$\Pi(s) = \prod_{p<z} (1-p^{-s}),$$

with $z = x^{(1-\vartheta)/2}$, and pick out terms with $x < n \leqslant x+x^\vartheta$. The right hand side produces $\sum_k k^{-1}(-1)^{k-1} S_k(\mathscr{A}, z)$, say, while the left hand side is essentially $\pi(x+x^\vartheta)-\pi(x)$. If one applies the Buchstab iteration to $S_k(\mathscr{A}, z)$ rather than $S(\mathscr{A}, z)$ one always has $P_i \leqslant x^{(1-\vartheta)/2}$, so that Vaughan's identity is unnecessary.

If $\mathscr{D}$ is the set of d for which $\sum_d' \# \mathscr{A}_d$ can be estimated accurately, the question of bounding $S(\mathscr{A}, z)$ from below may be formulated as a general sieve problem: maximize

$$\sum_{d \in \mathscr{D}} \lambda_d \# \mathscr{A}_d$$

subject to

$$\sum_{d|n, d \in \mathscr{D}} \lambda_d \leqslant \begin{cases} 1, & n = 1, \\ 0, & n > 1. \end{cases}$$

Since $\mathscr{D}$ is difficult to identify, no individual case of this sieve problem has been solved. None the less, for $\frac{4}{7} \leqslant \vartheta \leqslant \frac{7}{12}$, some fairly sharp bounds for $\pi(x+x^\vartheta)-\pi(x)$ are available. Specifically, when $\frac{4}{7} \leqslant \vartheta \leqslant \frac{7}{12}$, one can give asymptotic formulae for

$$S(\mathscr{A}, x^{\frac{1}{2}}) + \frac{1}{6} \# \{(p_1, \ldots, p_6);\quad p_i \leqslant x^{2(1-\vartheta)/5},\ p_1 \ldots p_6 \in \mathscr{A}\} \tag{5}$$

and

$$S(\mathscr{A}, x^{\frac{1}{2}}) - \frac{1}{3} \# \{(p_1, \ldots, p_5);$$
$$p_1, \ldots, p_4 \leqslant x^{2(1-\vartheta)/5},\quad p_5 \leqslant x^{4(1-\vartheta)/5},\quad p_1 \ldots p_5 \in \mathscr{A}\}. \tag{6}$$

These lead to

$$\frac{C_1(\vartheta)x^\vartheta}{\log x} \lesssim \pi(x+x^\vartheta)-\pi(x) \lesssim \frac{C_2(\vartheta)x^\vartheta}{\log x}$$

with

$$C_1(\vartheta) = 1-\frac{7^5}{288}\left(\frac{7}{12}-\vartheta\right)^4, \quad C_2(\vartheta) = 1+\frac{14^6}{1440}\left(\frac{7}{12}-\vartheta\right)^5.$$

In particular $C_1(\frac{4}{7}) \doteqdot 0.999980$, $C_2(\frac{4}{7}) \doteqdot 1.00000125$. The expression (5) arises from (4), taking $z = x^{1/7}$. For $k \geqslant 7$ we have $S_k(\mathscr{A}, z) = 0$, and for $k \leqslant 5$ we can give asymptotic formulae for $S_k(\mathscr{A}, z)$. There is also part of $S_6(\mathscr{A}, z)$ that can be dealt with, and the expression (5) reflects what is then left. For (6) the argument also starts with (4) but is more complicated. These ideas also yield new information at $\vartheta = \frac{7}{12}$. We have

$$\pi(x+x^{7/12})-\pi(x) = \frac{x^{7/12}}{\log x}\left\{1+O\left(\left(\frac{\log\log x}{\log x}\right)^4\right)\right\}.$$

A second application of this circle of ideas concerns intervals $(x, x+x^\vartheta]$ which "almost always" contain a prime. By this we mean that the set of $x \leqslant X$ for which the interval does not contain a prime has measure $o(X)$ as $X\to\infty$. It follows from Huxley's work [4] that if $\vartheta > \frac{1}{6}$ then there will almost always be asymptotically $x^\vartheta(\log x)^{-1}$ primes in the interval. Using sieve methods combined with ideas from the theory of zero density estimates it is now known that for $\vartheta > \frac{1}{12}$ the interval $(x, x+x^\vartheta]$ almost always contains $\geqslant (0\cdot 15)x^\vartheta(\log x)^{-1}$ primes. Let us see how this comes about.

Let $\mathscr{A}_k$ be as before. From the Buchstab formula we have

$$\begin{aligned} S(\mathscr{A}, x^{\frac{1}{2}}) &= S(\mathscr{A}, x^\varphi) - \sum_{x^\varphi < p \leqslant x^{\frac{1}{2}}} S(\mathscr{A}_p, p) \\ &= S(\mathscr{A}, x^\varphi) - \sum_{x^\varphi < p \leqslant x^{\frac{1}{2}}} S(\mathscr{A}_p, x^\varphi) + \sum_{x^\varphi < q < p \leqslant x^{\frac{1}{2}}} S(\mathscr{A}_{pq}, q). \qquad (7) \end{aligned}$$

If $\vartheta > \frac{1}{12}$ and $\varphi = \frac{5}{27}$ then one can give asymptotic formulae for the first two terms of (7) for almost all x, and the third term may be bounded below by zero. One can then conclude that

$$\pi(x+x^\vartheta)-\pi(x) \geqslant \frac{C(\vartheta)x^\vartheta}{\log x}$$

for almost all x. Unfortunately $C(\frac{1}{12}) \doteqdot -0.007$. It is therefore necessary to examine the discarded terms $S(\mathscr{A}_{pq}, q)$ to see what one can salvage. In fact there are substantial ranges of p, q that can be dealt with satisfactorily, to produce a saving of at least 0.16 in $C(\frac{1}{12})$.

Thus far I have not mentioned how the asymptotic formulae referred to may be obtained. I shall illustrate the process by considering

$$\#\{(p_1, \ldots, p_k); P_i < p_i \leqslant 2P_i, p_1 \ldots p_k \in \mathscr{A}\} = N,$$

say. We write

$$f_n(t) = \sum_{P_n < p \leqslant 2P_n} p^{-it},$$

and use the fact that for $y > 0$, $y \neq 1$ we have

$$\frac{1}{2\pi i} \int_{-T}^{T} (y^{it} - 1) \frac{dt}{t} = E(y) + O\left(\frac{1}{T|\log y|}\right), \tag{8}$$

where $E(y) = \frac{1}{2}$ $(y > 1)$ or $= -\frac{1}{2}(y < 1)$. Then

$$N = \frac{1}{2\pi i} \int_{-T}^{T} \{(x + x^{\theta})^{it} - x^{it}\} f_1(t) \ldots f_k(t) \frac{dt}{t} + O\left(\frac{x(\log x)}{T}\right).$$

Here we choose $T = x^{1-\theta}(\log x)^{2k}$, so that the error term becomes negligible. In the range $|t| \leqslant (\log x)^{2k}$ we can replace $(x+x^{\theta})^{it} - x^{it}$ by

$$\frac{x^{\theta}}{x(\log x)^{-4k}} \left\{\left(x + \frac{x}{(\log x)^{4k}}\right)^{it} - x^{it}\right\}$$

with negligible error. Then, using (8) in the reverse direction, we see that the range $|t| \leqslant (\log x)^{2k}$ relates N to the number of solutions of $p_1 \ldots p_k \in (x, x + x(\log x)^{-4k}]$. The latter may be calculated via the prime number theorem, and produces the main term of the asymptotic formula for N. There remains the interval $(\log x)^{2k} < |t| \leqslant T$, and it is here that the size and number of the P_i is important. Since

$$(x+x^{\theta})^{it} - x^{it} \ll |t| x^{\theta-1},$$

the relevent contribution is

$$\ll x^{\theta-1} \int_{(\log x)^{2k}}^{T} |f_1(t) \ldots f_k(t)| \, dt, \tag{9}$$

which we wish to be $O\big(x(\log x)^{-2k}\big)$, say. We can now use two tools from the theory of zero density estimates, namely the mean value theorem for Dirichlet polynomials and the Halász lemma. These may be applied to a number of different products of the f_i, producing various sets of conditions on the P_i under which (9) is sufficiently small. A further discussion of this may be found in Heath-Brown [2].

In conclusion I would like to mention one other problem where sieve methods have been successfully applied, namely Diophantine approximation with primes. Let α be irrational and let β be any real number. Vinogradov showed that if $\varepsilon > 0$ then

$$\|\alpha p+\beta\| \leqslant p^{\varepsilon-1/5} \tag{10}$$

for infinitely many primes. (Here $\|x\|$ is the distance from x to the nearest integer.) This was improved by Vaughan who showed that one can take $p^{-\frac{1}{4}}(\log p)^8$ on the right. In both cases the number of relevent primes in the appropriate interval was estimated asymptotically. By applying a sieve method in conjunction with Vaughan's estimates, Harman [1] has recently shown that one can take $p^{-3/10}$ on the right of (10). Here one uses bounds for exponential sums as opposed to the Dirichlet polynomial techniques of the previous problems. The bound $p^{-\frac{1}{3}}$ is beyond the scope of these methods.

Evidently there are many more potential applications of sieves in locating primes, and much work remains to be done. For example, it is not yet clear that the bound $p_{n+1}-p_n \leqslant p_n^{\frac{1}{2}+\varepsilon}$ is out of reach. I feel that the significance of these ideas is not yet fully appreciated, nor is their power fully exploited.

References

[1] Harman G., On the Distribution of αp Modulo 1, *J. London Math. Soc.* (2) **27** (1983), pp. 9–18.

[2] Heath-Brown D. R., Prime Numbers in Short Intervals and a Generalized Vaughan Identity, *Can. J. Math.* **34** (1982), pp. 1365–1377.

[3] Heath-Brown D. R. and Iwaniec H., On the Difference Between Consecutive Primes, *Inventiones Math.* **55** (1979), pp. 49–69.

[4] Huxley M. N., On the Difference Between Consecutive Primes, *Inventiones Math.* **15** (1972), pp. 164–170.

[5] Iwaniec H., *Primes in Short Intervals*, to appear.

[6] Iwaniec H. and Jutila M., Primes in Short Intervals, *Arkiv för Mat.* **17** (1979), pp. 167–176.

[7] Pintz J., *On Primes in Short Intervals* I, to appear.

MAGDALEN COLLEGE,
OXFORD OX1 4AU,
ENGLAND

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

D. W. MASSER

Zero Estimates on Group Varieties

1. Introduction

The aim of this article is to give an account of a class of estimates that have in recent years been found useful in the theory of transcendental numbers. As our space is limited, and the estimates are quite technical in nature, we can provide no hints of proofs, and we prefer to illustrate the results with examples rather than supply full statements. We also mention some of the applications.

One basic method in transcendence theory can be described in general terms as follows. Suppose it is desired to establish a certain proposition. Usually it will be possible to associate with the problem a collection of meromorphic functions $f_1, \ldots, f_n$ in such a way that assuming the falsity of the proposition imposes strong arithmetical conditions on the functions. These enable us to construct a polynomial $P = P(x_1, \ldots, x_n)$ such that the function $\phi = P(f_1, \ldots, f_n)$ is not identically zero but has a large number of zeroes. By means of analytic interpolation arguments it can then be deduced that ϕ takes many small values, and by using once more the arithmetical conditions we conclude that ϕ has even more zeroes than initially arranged. But often merely knowing upper bounds for the degree of P leads to upper bounds for the number of zeroes of ϕ, and if we are lucky these yield the contradiction that proves the original proposition.

In this article we shall limit attention to the last stage of this argument, as it is here that the most interesting recent progress has taken place. A simple example will suffice to illustrate the sort of estimates involved. To show that w and e^w cannot both be algebraic for $w \neq 0$, suppose they are. Then we end up with a non-zero polynomial $P = P(x, y)$, of degree at most $L \geqslant 1$ in x and of degree at most $M \geqslant 1$ in y, such that the function

$$\phi(z) = P(z, e^z) \tag{1}$$

has a zero of order at least T at the points $0, w, \ldots, Sw$ for some S. It is not hard to prove (see for example [7]) that if

$$TS > 4LM, \qquad T > 2M, \tag{2}$$

then in fact $P = 0$. Thus if we can make T and S so large as to satisfy (2), we obtain our desired contradiction.

We shall refer to the above type of statement as a zero estimate. In the last few years such results have been systematized and put into a general context, whereas before this, many ingenious ad hoc arguments, often involving manipulation of determinants or Kummer theory, were found necessary to finish the proofs.

The essential tool in the recent developments has been commutative algebra. This was introduced in the fundamental paper of Nesterenko [25]. His results were designed to apply to Siegel E-functions, and so they treat the case when $f_1, \ldots, f_n$ are functions of a single complex variable satisfying linear differential equations over $C(z)$. Later on Brownawell and the author [8], following the broad outline of [25] but changing the details somewhat, were able to treat non-linear differential equations over C (see also Brownawell [6]). All this work is directed essentially at bounding the order of a single zero of ϕ.

2. Bounds without multiplicities

If instead we want to estimate the number of zeroes of ϕ on a given set counted without multiplicity, then provided the set has suitable translation properties, it is appropriate to consider translation formulae rather than differential equations. A general context (though not the most general) where such formulae are available is that of group varieties. In this case $f_1, \ldots, f_n$ are more or less the coordinates of the associated exponential map, so it is possible to work directly on the group variety without mentioning meromorphic functions. In the paper [19] of Wüstholz and the author the following situation is studied.

Let G be a commutative group variety of dimension $n \geqslant 1$, assumed embedded in projective space P_N of $N \geqslant 1$ dimensions, and let Γ be a finitely generated subgroup of G. For each integer r with $1 \leqslant r \leqslant n$ define

$$p_r = p_r(\Gamma, G) = \min_H \operatorname{rank}(\Gamma/\Gamma \cap H),$$

where the minimum is taken over all algebraic subgroups H of G of dimension $n-r$ (with $p_r = \operatorname{rank}\Gamma$ if no such H exists). Let

$$\mu = \mu(\Gamma, G) = \min_{1 \leqslant r \leqslant n} (p_r/r).$$

Let $\gamma_1, \ldots, \gamma_m$ be generators of Γ and for $S \geqslant 0$ write $\Gamma(S)$ for the set of linear combinations $s_1\gamma_1 + \ldots + s_m\gamma_m$ as $s_1, \ldots, s_m$ run over all integers with $0 \leqslant s_1, \ldots, s_m \leqslant S$. The main theorem of [19] states that there is a constant c, depending only on G and the embedding, with the following property. If $P = P(X_0, \ldots, X_N)$ is a homogeneous polynomial of degree $D \geqslant 1$ vanishing on $\Gamma(S)$ with

$$(S/n)^{\mu} > cD \tag{3}$$

then P vanishes identically on G. The exponent μ in (3) can be shown to be best possible for any given G and Γ.

We should note here that Moreau [23] has given a much shorter version of the proof of this result, using geometric arguments in place of the commutative algebra (see also [24]). But it is not clear if his work extends to multiplicities.

When $G = A$ is an abelian variety defined over $\bar{Q}$, this result has been used in [18] to prove certain lower bounds for the Néron–Tate height on the set of points of A defined over $\bar{Q}$. When $G = \boldsymbol{G}_m^n$ is the n-fold product of the multiplicative group $\boldsymbol{G}_m$, it has been used by Waldschmidt [37] to prove some generalizations of certain transcendence results on numbers of the form e^{uv} (see also [39]). In particular these answer questions of Serre and Weil on characters (see [38]), and they have interesting consequences for p-adic regulators (see [40] and [41]).

We illustrate the above zero estimate when G is the product of $\boldsymbol{G}_m$ by the additive group $\boldsymbol{G}_a$. First of all embed G in $\boldsymbol{P}_2$ as the set of points (x_0, x_1, x_2) with $x_0 x_2 \neq 0$. If (y_0, y_1, y_2) is another such point, their group sum is $(x_0y_0, x_0y_1 + x_1y_0, x_2y_2)$. The constant c of [19] turns out to be 1. If Γ is generated by the point $\gamma = (1, w, e^w)$, it is not difficult to see that $\mu(\Gamma, G) = 1/2$ provided $w/2\pi i$ is irrational. So in this case let $P(x, y)$ be a polynomial of total degree at most $D \geqslant 1$ such that the function (1) vanishes at the points $0, w, \ldots, Sw$ for

$$S > 2D^2.$$

Then the homogeneous polynomial

$$P(X_0, X_1, X_2) = X_0^D P(X_1/X_0, X_2/X_0) \tag{4}$$

vanishes on $\Gamma(S)$, and we conclude that $P = 0$.

Similarly by considering $G = \boldsymbol{G}_m^n$ in $\boldsymbol{P}_n$ we can obtain zero estimates for $\phi(z) = P(e^{u_1 z}, \ldots, e^{u_n z})$. Sometimes these can be proved by means of Tijdeman's powerful analytic methods [33], [35], which by virtue of their simplicity and elegance have very often been found indispensable for problems involving exponential polynomials. But the strength of the algebraic method lies in its generality; for example, if $\wp(z)$ is a Weierstrass elliptic function we can obtain zero estimates for $P(\wp(u_1 z), \ldots, \wp(u_n z))$ simply by considering $G = E^n$ for an appropriate elliptic curve E. Such estimates were used in [20] to establish elliptic analogues of some algebraic independence results for numbers of the form e^{uv}. Thus if $\wp(z)$ has algebraic invariants and no complex multiplication, and $u_1, \ldots, u_n$ as well as $v_1, \ldots, v_m$ are complex numbers linearly independent over $\mathbf{Q}$, then provided

$$mn \geqslant 2m + 4n \tag{5}$$

at least two of the numbers

$$\wp(u_i v_j) \qquad (1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m) \tag{6}$$

are algebraically independent. The exponential analogue of this had been obtained independently by Brownawell [5], Smelev [32], Waldschmidt [36] and Wallisser (see also Tijdeman [34]).

By suitably refining these zero estimates similar results can be given for larger transcendence degrees. Thus Wüstholz and the author show in [21] that, with mild linear independence measures on $u_1, \ldots, u_n$ and $v_1, \ldots, v_m$, the condition

$$mn \geqslant 2^{k+1}(k+7)(m+2n)$$

is enough to ensure that at least k of the numbers (6) are algebraically independent. This is an analogue of a result of Chudnovsky [9] (see also Warkentin [42], Reyssat [31], Philippon [27], Endell [11] and Nesterenko [26]).

3. Multiplicities in a single direction

So far we have considered two types of zero estimate: those for a single high order zero, and those for many zeroes without multiplicity. In [22] these are combined (see also the announcements in [19]). For functions of a single complex variable the natural concept is that of a one-parameter subgroup of a group variety G. This is a non-zero analytic group homomorphism Λ from $\mathbf{C}$ to G. If as before G is embedded in $\mathbf{P}_N$ and $P = P(X_0, \ldots, X_N)$ is a homogeneous polynomial it is easy to define for any g in G the order of vanishing of P at g along Λ. With the notation above, one of the results of [22] can be stated as follows. There is a constant c, depending only on G and the embedding, with the following property. If P has degree $D \geqslant 1$ and vanishes at each point of $\Gamma(S)$ to order at least T along A with

$$T(S/n)^{p_r} > cD^r \qquad (1 \leqslant r \leqslant n), \tag{7}$$

$$(S/n)^{p_r} > cD^{r-1} \qquad (1 \leqslant r < n), \tag{8}$$

then P vanishes on $g + A(C)$ for some g in G. The conditions (8) are not natural and they reflect a technical difficulty in the proof. Note also that the conditions (7) for $T = 1$ reduce essentially to (3).

As an example, take $G = G_a \times G_m$ in P_2 as above, and put $A(z) = (1, z, e^z)$. For a homogeneous polynomial $P = P(X_0, X_1, X_2)$ and a point $g = (x_0, x_1, x_2)$ on G the order of vanishing of P at g along A turns out to be simply the order of zero at $z = 0$ of the function $P(x_0, x_0 z + x_1, x_2 e^z)$. Thus if the function (1) has a zero of order t at $z = sw$ for some integer s, then the polynomial P defined by (4) vanishes at $s\gamma$ to order t along A, where $s\gamma = (1, sw, e^{sw})$. Hence we obtain multiplicity estimates for such functions.

The elliptic analogues lead to corresponding generalizations of the above algebraic independence results. For example, the condition (5) may be relaxed to $mn \geqslant 2m + 2n$ as long as we adjoin the numbers $u_1, \ldots, u_n$ to (6). If $\wp(z)$ has complex multiplication over a quadratic field k, the same result holds provided now $u_1, \ldots, u_n$ are linearly independent over k. A corollary is that if β is cubic over k and $u \neq 0$ is such that $\wp(u)$ is algebraic, then $\wp(\beta u)$ and $\wp(\beta^2 u)$ are algebraically independent. This is the elliptic analogue of Gelfond's result on α^β and α^{β^2}.

The work of [22] includes some other refinements. Firstly the results are stated for multihomogeneous polynomials; this allows us to have different degrees in different variables, as in the original example (1). Secondly, a more delicate measure of the distribution of Γ with respect to algebraic subgroups is introduced; for example, if Γ consists only of torsion points then $p_1 = \ldots = p_n = \mu = 0$ and so the conditions (3), (7), (8) are too restrictive. This last refinement is especially valuable because Wüstholz has shown how in certain circumstances it can be used to eliminate the technical conditions (8). Furthermore in [43] he deduces the following remarkable consequence for an elliptic function $\wp(z)$ with algebraic invariants and complex multiplication over k; if $\alpha_1, \ldots, \alpha_n$ are algebraic numbers linearly independent over k, then $\wp(\alpha_1), \ldots, \wp(\alpha_n)$ are algebraically independent. This is of course the analogue of the celebrated Lindemann–Weierstrass theorem of 1885. It had previously been proved by Chudnovsky [10] for $n = 1, 2, 3$.

We should also mention that at about the same time Philippon [29] found a different though related approach to this result, still using zero estimates. Furthermore in [28] (see also [30]) he obtains slightly weaker results for the numbers $\wp(u), \wp(\beta u), \ldots, \wp(\beta^{d-1}u)$; nevertheless these are still far in advance of the known exponential analogues.

4. Arbitrary multiplicities

We end by discussing briefly the most exciting recent developments; namely, the extension by Wüstholz [44] of the multiplicity estimates to several variables. If G is a group variety of dimension n, and d is a fixed integer with $1 \leqslant d \leqslant n$, a d-parameter subgroup of G is an analytic group homomorphism Λ from $\boldsymbol{C}^d$ to G whose Jacobian is not identically zero. The main result of [44] then directly generalizes the earlier multiplicity estimate to such Λ.

Now when this result is interpreted in terms of meromorphic functions, it involves partial differentiation with respect to d variables. The case $d = n-1$ is especially important, since it corresponds to Baker's method in transcendence theory. For example, with $G = \boldsymbol{G}_m^n$ embedded in $\boldsymbol{P}_n$ as the set of points $(x_0, \ldots, x_n)$ with $x_0 \ldots x_n \neq 0$, let $\beta_1, \ldots, \beta_{n-1}$ be algebraic numbers with $1, \beta_1, \ldots, \beta_{n-1}$ linearly independent over $\boldsymbol{Q}$, and take $\Lambda(z_1, \ldots, z_{n-1}) = (1, e^{z_1}, \ldots, e^{z_{n-1}}, e^{\beta_1 z_1 + \ldots + \beta_{n-1} z_{n-1}})$. Let $\alpha_1, \ldots, \alpha_{n-1}$ be non-zero algebraic numbers with logarithms $l_1, \ldots, l_{n-1}$ not all zero. If $\alpha_1^{\beta_1} \ldots \alpha_{n-1}^{\beta_{n-1}}$ is algebraic, Baker's method [1] yields a non-zero polynomial $P = P(x_1, \ldots, x_n)$, of total degree at most $D \geqslant 1$, such that the function

$$\phi(z_1, \ldots, z_{n-1}) = P(e^{z_1}, \ldots, e^{z_{n-1}}, e^{\beta_1 z_1 + \ldots + \beta_{n-1} z_{n-1}})$$

has zeroes of order at least T at the points

$$(z_1, \ldots, z_{n-1}) = s(l_1, \ldots, l_{n-1}) \quad (0 \leqslant s \leqslant S)$$

with

$$T^{n-1}S > cD^n, \quad T > cD$$

for any large constant c. The main result of [44] then implies that $P = 0$. Hence $\alpha_1^{\beta_1} \ldots \alpha_{n-1}^{\beta_{n-1}}$ is transcendental.

But now the generality of the algebraic approach means that the analogues for any commutative group variety can be proved in the same way. Some of these are worked out in [45] (see also [46]); for example, if $\wp(z)$ has algebraic invariants and $\alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{n-1}$ are as above, it is shown that $\alpha_1^{\beta_1} \ldots \alpha_{n-1}^{\beta_{n-1}} e^{\gamma_1 u_1} \ldots e^{\gamma_m u_m}$ is transcendental for any algebraic $\gamma_1, \ldots, \gamma_m$ and any complex numbers $u_1, \ldots, u_m$ such that $\wp(u_1), \ldots$ $\ldots, \wp(u_m)$ are algebraic. In terms of linear forms, this result mixes ordinary logarithms with elliptic logarithms.

By refining his zero estimates to deal with torsion points, Wüstholz settles two other linear forms problems that seemed hopeless a few years ago. The first of these was raised by Baker [1]. Let $\wp_1(z), \ldots, \wp_n(z)$ have algebraic invariants, and let $\zeta_1(z), \ldots, \zeta_n(z)$ be the corresponding Weierstrass zeta functions. Pick periods $\omega_1, \ldots, \omega_n$ of $\wp_1(z), \ldots, \wp_n(z)$ respectively, and put

$$\eta_i = \zeta_i(z+\omega_i) - \zeta_i(z) \qquad (1 \leqslant i \leqslant n).$$

In [47] it is shown that any linear form in $\omega_1, \ldots, \omega_n, \eta_1, \ldots, \eta_n$ and $2\pi i$ with algebraic coefficients is either zero or transcendental. This was previously known only for $n = 1, 2$. Furthermore Wüstholz determines when such a linear form can vanish; this had not even been settled for $n = 2$ (see [16] for a more detailed history). The proofs involve the group $G = G' \times \boldsymbol{G}_m$, where G' is an extension of a product of elliptic curves by $\boldsymbol{G}_a$.

The second problem was raised by Bertrand [2]. Take now a single pair of functions $\wp(z)$, $\zeta(z)$ and numbers ω, η as above. Write $\lambda(z) = \omega\zeta(z) - \eta z$, and let $u_1, \ldots, u_n$ be such that $\wp(u_1), \ldots, \wp(u_n)$ are algebraic. Then in [48] it is shown that any linear form in $\omega, \eta, \lambda(u_1), \ldots, \lambda(u_n)$ with algebraic coefficients is either zero or transcendental, and again the two possibilities can be distinguished. For $n = 1, 2$ these results had been proved by Laurent [14]; see also the very interesting related work of Bertrand [3] and their joint article [4]. This time the group G is an extension of an elliptic curve by $\boldsymbol{G}_m^n \times \boldsymbol{G}_a$. As a consequence, one can now decide whether a period of an arbitrary differential on an elliptic curve is transcendental or not, provided both are defined over $\overline{\boldsymbol{Q}}$.

Finally the techniques of [44] also lead to quantitative results good enough for applications to diophantine problems. For example, when combined with a method of Lang [13], they provide a new proof of Siegel's theorem for arbitrary curves. Furthermore, the only ineffective step in the proof consists of the determination of a basis of the corresponding Mordell–Weil group. Previously this approach could only be made to work in the case of complex multiplication [15].

References

[1] Baker A., *Transcendental Number Theory*, Cambridge, 1975.

[2] Bertrand D., Abelian Functions and Transcendence, *Journées Arithmétiques 1980* (ed. J. V. Armitage), London Math. Soc. Lecture Notes **56**, Cambridge, 1982, pp. 1–10.

[3] Bertrand D., Valeurs de fonctions thêta et hauteurs p-adiques, *Sém. Theorie Nombres Paris (Sém. Delange–Pisot–Poitou) 1980/81, Progress in Math.* **22**, Birkhäuser, Boston–Basel–Stuttgart, 1982, pp. 1–11.

[4] Bertrand D. and Laurent M., Propriétés de transcendance de nombres liés aux fonctions thêta, *C. R. Acad. Sci. Paris* **292** (1981), pp. 747–749.

[5] Brownawell W. D., The Algebraic Independence of Certain Values of the Exponential Function, *K. Norske Vidensk. Selsk. Skr.* **23** (1972), 5p.

[6] Brownawell, W. D., Zero Estimates for Solutions of Differential Equations, *Approximations diophantiennes et nombres transcendants, Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Brikhäuser, Boston–Basel–Stuttgart, 1983, pp. 67–94.

[7] Brownawell W. D. and Masser D. W., Multiplicity Estimates for Analytic Functions I, *J. reine angew. Math.* **314** (1980), pp. 200–216.

[8] Brownawell W. D. and Masser D. W., Multiplicity Estimates for Analytic Functions II, *Duke Math. J.* **47** (1980), pp. 273–295.

[9] Chudnovsky G. V., *Some Analytic Methods in the Theory of Transcendental Numbers*, Inst. of Math., Ukr. Acad. Sci., Preprint IM 74-8 (48 p.) and IM 74-9 (52 p.), Kiev, 1974.

[10] Chudnovsky G. V., Algebraic Independence of the Values of Elliptic Functions at Algebraic Points; Elliptic Analogue of tho Lindemann–Weierstrass Theorem, *Inventiones Math.* **61** (1980), pp. 267–290.

[11] Endell R., *Zur algebraischen Unabhängigkeit gewisser Werte der Exponentialfunktion (nach Chudnovsky)*, Diplomarbeit, Düsseldorf, 1981.

[12] Gelfond A. O., *Transcendental and Algebraic Numbers*, Dover, New York, 1960.

[13] Lang S., Diophantine Approximation on Toruses, *Amer. J. Math.* **86** (1964), pp. 521–533.

[14] Laurent M., Transcendance de périodes d'intégrales elliptiques II, *J. reine angew. Math.* **333** (1982), pp. 144–161.

[15] Masser D. W., Linear Forms in Algebraic Points of Abelian Functions III, *Proc. London Math. Soc.* **33** (1976), pp. 549–564.

[16] Masser D. W., Some Vector Spaces Associated with Two Elliptic Functions, *Transcendence Theory: Advances and Applications* (eds. A.Baker and D.W. Masser), Academic Press, London, 1977, pp. 101–119.

[17] Masser D. W., On Polynomials and Exponential Polynomials in Several Complex Variables, *Inventiones Math.* **63** (1981), pp. 81–95.

[18] Masser D. W., Small Values of the Quadratic Part of the Néron–Tate Height, *Sém. Theorie Nombres Paris (Sém. Delange–Pisot–Poitou) 1979/80, Progress in Math.* **12**, Birkhäuser, Boston–Basel–Stuttgart, 1981, pp. 213–222; proofs to appear in *Compositio Math.*

[19] Masser D. W. and Wüstholz G., Zero Estimates on Group Varieties I, *Inventiones Math.* **64** (1981), pp. 489–516.

[20] Masser D. W. and Wüstholz G., Algebraic Independence Properties of Values of Elliptic Functions, *Journées Arithmétiques 1980* (ed. J. V. Armitage), London Math. Soc. Lecture Notes **56**, Cambridge, 1982, pp. 360–363.
[21] Masser D. W. and Wüstholz G., Fields of Large Transcendence Degree Generated by Values of Elliptic Functions, *Inventiones Math.* **72** (1983), pp. 407–464.
[22] Masser D. W. and Wüstholz G., Zero Estimates on Group Varieties II.
[23] Moreau J. C., Démonstrations géométriques de lemmes de zeros I, *Sém. Theorie Nombres Paris* (*Sém. Delange–Pisot–Poitou*) *1981/82*, *Progress in Math.*, Birkhäuser, Boston–Basel–Stuttgart.
[24] Moreau J. C., Démonstrations géométriques de lemmes de zeros II, *Approximations diophantiennes et nombres transcendants*, *Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Birkhäuser, Boston–Basel–Stuttgart, 1983, pp. 191–197.
[25] Nesterenko J. V., Estimates for the Orders of Zeroes of Functions of a Certain Class and Their Applications to the Theory of Transcendental Numbers, *Izv. Akad. Nauk. USSR, Ser. Mat.* **41** (1977), pp. 253–284; *Math. USSR Izv.* **11** (1977), pp. 239–270.
[26] Nesterenko J. V., On the Algebraic Independence of Algebraic Numbers to Algebraic Powers, *Approximations diophantiennes et nombres transcendants*, *Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Birkhäuser, Boston–Basel–Stuttgart, 1983, pp. 199–220.
[27] Philippon P., Indépendance algébrique de valeurs de fonctions exponentielles p-adiques, *J. reine angew. Math.* **329** (1981), pp. 42–51.
[28] Philippon P., Variétés abéliennes et indépendance algébrique I, *Inventiones Math.* **70** (1983), pp. 289–318.
[29] Philippon P., Variétés abéliennes et indépendance algébrique II, *Inventiones Math.* **72** (1983), pp. 389–405.
[30] Philippon P., Sous-groupes à n paramètres et indépendance algébrique, *Approximations diophantiennes et nombres transcendants*, *Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Birkhäuser, Boston–Basel–Stuttgart, 1983, pp. 221–234.
[31] Reyssat E., Un critère d'indépendance algébrique, *J. reine angew. Math.* **329** (1981), pp. 66–81.
[32] Smelev A. A., On the Method of A. O. Gelfond in the Theory of Transcendental Numbers, *Mat. Zametki* **10** (1971), pp. 415–426; *Math. Notes* **10** (1971), pp. 672–678.
[33] Tijdeman R., On the Number of Zeroes of General Exponential Polynomials, *Indagationes Math.* **33** (1971), pp. 1–7.
[34] Tijdeman R., On the Algebraic Independence of Certain Numbers, *Indagationes Math.* **33** (1971), pp. 146–162.
[35] Tijdeman R., An Auxiliary Result in the Theory of Transcendental Numbers, *J. Number Th.* **5** (1973), pp. 80–94.
[36] Waldschmidt M., Indépendance algébrique des valeurs de la fonction exponentielle, *Bull. Soc. Math. France* **99** (1971), pp. 285–304.
[37] Waldschmidt M., Transcendance et exponentielles en plusieurs variables, *Inventiones Math.* **63** (1981), pp. 97–127.
[38] Waldschmidt M., Sur certains caractères du groupe des classes d'idèles d'un corps de nombres, *Sém. Theorie Nombres Paris* (*Sém. Delange–Pisot–Poitou*)

1980/81, Progress in Math. **22**, Birkhäuser, Boston–Basel–Stuttgart, 1982, pp. 323–335.

[39] Waldschmidt M., Sous-groupes analytiques de groupes algébriques, *Ann. of Math.* **117** (1983), pp. 627–657.

[40] Waldschmidt M., A Lower Bound for the p-adic Rank of the Units of an Algebraic Number Field, Coll. Number Theory, Janos Bolyai Math. Soc. Budapest, July 1981; *Topics in Classical Number Theory* **34**.

[41] Waldschmidt M., Dépendance de logarithmes dans les groupes algébriques, *Approximations diophantiennes et nombres transcendants, Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Birkhäuser, Boston–Basel–Stuttgart, 1983, pp. 289–328.

[42] Warkentin P., *Algebraische Unabhängigkeit gewisser p-adischer Zahlen*, Diplomarbeit, Freiburg, 1978.

[43] Wüstholz G., Über das abelsche Analogon des Lindemannschen Satzes, *Inventiones Math.* **72** (1983), pp. 363–388.

[44] Wüstholz G., *Multiplizitätsabschätzungen auf algebraischen Gruppen.*

[45] Wüstholz G., Some Remarks on a Conjecture of M. Waldschmidt, *Approximations diophantiennes et nombres transcendants, Luminy 1982* (eds. D. Bertrand and M. Waldschmidt), *Progress in Math.* **31**, Birkhäuser, Boston–Basel–Stuttgart, 1983, pp. 329–336.

[46] Wüstholz G., *Algebraic Values of Analytic Homomorphisms of Algebraic Groups*, Preprint.

[47] Wüstholz G., Zum Periodenproblem, to appear in *Inventiones Math.*

[48] Wüstholz G., Transzendenzeigenschaften von Perioden elliptischer Integrale, to appear in *J. reine angew. Math.*

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NOTTINGHAM
UNIVERSITY PARK
NOTTINGHAM, U.K.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

KENNETH A. RIBET*

Congruence Relations between Modular Forms

1. This article is concerned with the notion of *congruence primes* in the theory of modular forms, as in the work of Doi and Hida [1], Doi and Ohta [2], and Hida [3], [4], [5]. Our main aim is to point out how the explicit calculation of such primes, in a particular example involving forms of weight 2, leads to a non-trivial problem concerning finite subgroups of Jacobians of modular curves.

Let $k \geqslant 2$ and $N \geqslant 1$ be integers, and take S to be either the complex vector space of holomorphic modular forms or the vector space of holomorphic cusp forms of weight k on one of the classical modular groups $\Gamma_0(N)$ or $\Gamma_1(N)$. We denote by $S(\mathbf{Z})$ the lattice of forms in S with integral q-expansion, and by T_n (for $n \geqslant 1$) the n^{th} Hecke operator on S.

Suppose that we are given a direct sum decomposition

$$S = X \oplus Y \tag{1.1}$$

in which X and Y are both stable under the T_n and both generated by their intersections with $S(\mathbf{Z})$. A prime number p is a *congruence prime* relative to this decomposition if there is a non-trivial mod p congruence linking X to Y: there exist

$$f \in X \cap S(\mathbf{Z}), \quad g \in Y \cap S(\mathbf{Z})$$

such that

$$f \equiv g \bmod pS(\mathbf{Z}), \quad f \not\equiv 0 \bmod pS(\mathbf{Z}).$$

For example, taking S to be the space of weight-k modular forms on $\mathbf{SL}(2, \mathbf{Z})$, we may choose X (resp. Y) to be the space of Eisenstein series (resp. cusp forms) in S. The congruence primes are those prime numbers which divide the numerator of the constant term of the normali-

* Research partially supported by the National Science Foundation (USA).

zed Eisenstein series of weight k, i.e., the fraction $B_k/2k$, where B_k is the k^{th} Bernoulli number. Thus congruence primes are irregular primes, and the congruence link between X and Y may be used in studying the arithmetic of cyclotomic fields. Doi has asked whether, more generally, one can characterize congruence primes and interpret the link between X and Y in terms of arithmetic.

In his articles citied above, Hida discussed these questions of Doi, the first quite generally, and the second in reference to cusp forms with complex multiplication. Especially, the articles [3] and [4], together with the author's [10], give an interpretation of congruence primes in terms of parabolic cohomology. Here we assume that S is a space of cusp forms and use the well known Shimura isomorphism to realize S as a certain parabolic cohomology group V constructed with real coefficients. Via this isomorphism, S is endowed with a second integral lattice $V(\mathbf{Z})$, the image in V of the analogous cohomology group made with integral coefficients. Replacing $S(\mathbf{Z})$ by $V(\mathbf{Z})$ in the definition of "congruence prime", we obtain the alternate notion of *cohomology congruence prime.*

THEOREM 1.2 ([4], [10]). *Every cohomology congruence prime is a congruence prime. Conversely, if p is a congruence prime not dividing $(k-1)!N$, then p is a cohomology congruence prime.*

This theorem shows that the two notions of congruence prime are essentially equivalent. On the other hand, one feels that the set of cohomology congruence primes may be precisely calculated in certain contexts. (For example, Hida showed in some cases how the cohomology congruence primes are the prime divisors of a rational integer which may be interpreted as the "algebraic part" of a determinant of periods of forms in X.)

2. To test this idea, we are going to work out an explicit example. Since it is more pleasing to consider congruences between eigenforms for the Hecke operators, rather than congruences between arbitrary forms, we begin by reviewing the notion of *primes of fusion.* These will be maximal ideals of the Hecke ring associated to S whose residue characteristics are precisely the congruence primes.

We let $\mathbf{T}$ be the subring of $\mathrm{End}(S)$ generated by the Hecke operators T_n acting on S, and we similarly define $\mathbf{T}_X$ and $\mathbf{T}_Y$ by replacing S by X and Y. Then $\mathbf{T}_X$ and $\mathbf{T}_Y$ are naturally quotients of $\mathbf{T}$, which in turn is a subring of the direct sum $\mathbf{T}_X \oplus \mathbf{T}_Y$. A *prime of fusion* is a prime ideal of $\mathbf{T}$ containing the conductor of the ring extension

$$\mathbf{T} \hookrightarrow \mathbf{T}_X \oplus \mathbf{T}_Y.$$

If $\mathscr{P}$ is such a prime, its image in $\mathbf{T}_X$ (resp. $\mathbf{T}_Y$) is a prime ideal $\mathscr{P}_X$ (resp. $\mathscr{P}_Y$) of $\mathbf{T}_X$ (resp. $\mathbf{T}_Y$). We again refer to $\mathscr{P}_X$ and $\mathscr{P}_Y$ as primes of fusion, and we note the canonical isomorphisms

$$\mathbf{T}_X/\mathscr{P}_X \simeq \mathbf{T}/\mathscr{P} \simeq \mathbf{T}_Y/\mathscr{P}_Y.$$

Especially, we view the isomorphism between extreme terms as a congruence between eigenvalues of the T_n on X and on Y (cf. [3], Th. 7.1).

Conversely, suppose that $f \in X$ and $g \in Y$ are eigenforms for the T_n, with eigenvalues a_n and b_n ($n \geqslant 1$) respectively. Let $\mathcal{O}$ be the ring of integers of the number field generated by all the a_n and b_n. Then there are unique homomorphisms

$$\phi_X\colon \mathbf{T}_X \to \mathcal{O}, \quad \phi_Y\colon \mathbf{T}_Y \to \mathcal{O}$$

such that $\phi_X(T_n) = a_n$ and $\phi_Y(T_n) = b_n$ for all n. Assume now that λ is a maximal ideal of O such that

$$a_n \equiv b_n \bmod \lambda$$

for all n. Then one sees immediately that $\mathscr{P}_X = \phi_X^{-1}(\lambda)$ and $\mathscr{P}_Y = \phi_Y^{-1}(\lambda)$ are primes of fusion in $\mathbf{T}_X$ and $\mathbf{T}_Y$. The corresponding ideal $\mathscr{P}$ of $\mathbf{T}$ is obtained by pulling back either $\mathscr{P}_X$ or $\mathscr{P}_Y$ to $\mathbf{T}$.

If $\mathscr{P}_X$ is an ideal of $\mathbf{T}_X$ which one suspects to be a prime of fusion, one can prove that $\mathscr{P}_X$ is indeed such a prime by exhibiting a $\mathbf{T}_X$-module Ω, whose support contains $\mathscr{P}_X$, which satisfies the following condition: if we view Ω as a $\mathbf{T}$-module via the natural surjection $\mathbf{T} \to \mathbf{T}_X$, the resulting homomorphism

$$\mathbf{T} \to \mathrm{End}(\Omega)$$

factors through the surjection $\mathbf{T} \to \mathbf{T}_Y$ as well. In Proposition 1.11 of [10], the author showed that a certain $\mathbf{T}$-module $\tilde{L}/L$ detects in this way all primes of fusion which do not divide the level N of the space S. In other words, one can find essentially all primes of fusion by calculating the support of this module.

We now come to the specific problem alluded to above. We will consider only weight 2 cusp forms (and especially newforms) on groups of the form $\Gamma_0(N)$. We first take a newform

$$f = \sum a_n q^n$$

on $\Gamma_0(N)$ and then consider a prime number M which is prime to N. Suppose that λ is a prime ideal in the ring of integers of a sufficiently large finite extension of $\mathbf{Q}$ in $\bar{\mathbf{Q}}$ whose residue characteristic l is prime to

MN. Suppose that

$$g = \sum b_n q^n$$

is a weight 2 newform of level divisible by M and dividing NM for which we have the congruence

$$a_p \equiv b_p \mod \lambda \tag{2.1}$$

for all prime numbers p in a set of primes of density 1. Then the mod λ representations of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$ defined by f and g have isomorphic semi-simplifications. By considering the restrictions of these representations to the decomposition group $\mathrm{Gal}(\bar{\mathbf{Q}}_M/\mathbf{Q}_M)$ for M in $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$, we obtain the congruence

$$a_M \equiv b_M(1+M) \mod \lambda.$$

Since b_M is either $+1$ or -1, we find

$$a_M^2 \equiv (1+M)^2 \mod \lambda. \tag{2.2}$$

Our problem is to determine whether or not the converse holds: if the congruence (2.2) is verified for a specific prime $M \nmid N$, need there be a form g satisfying (2.1)?

We can rephrase this problem in terms of primes of fusion by considering a suitable decomposition (1.1) of the space S of weight 2 cusp forms on $\Gamma_0(NM)$. Namely, we take X to be the subspace of old forms of S which is associated to $\Gamma_0(N)$, so that X is naturally isomorphic to the direct sum of two copies of the space of cusp forms of weight 2 on $\Gamma_0(N)$. We then take Y to be the orthogonal complement to X under the Petersson inner product on S. Thus Y is the intersection of the kernels of the two natural trace maps from S to the space of weight 2 cusp forms on $\Gamma_0(N)$. Our problem will be to show, under hypothesis (2.2), that a certain ideal $\mathscr{P}_X$ of $\mathbf{T}_X$ is a prime of fusion.

To define $\mathscr{P}_X$, our inclination would be to proceed as before, using f to define a homomorphism ϕ_X, and using ϕ_X to pull back λ to $\mathbf{T}_X$. A technical complication arises, however: f is no longer an eigenform for the Hecke operator T_M acting on X. To surmount this, assuming that we have

$$a_M \equiv \pm(1+M) \mod \lambda,$$

we introduce

$$f' = \sum a_n q^n \mp M \sum a_n q^{Mn} \in X.$$

Then f' is a mod λ eigenform for T_M with eigenvalue ± 1; it is also an eigenform for the T_n such that $(n, M) = 1$, with eigenvalue a_n. Thus f' defines a homomorphism $\mathbf{T}_X \to \mathbf{F}$, where $\mathbf{F}$ is the residue field of λ, such that the image of T_n for each $n \geqslant 1$ is the eigenvalue of T_n acting on f', modulo λ. Its kernel is a prime ideal $\mathscr{P}_X$ of $\mathbf{T}_X$; pulling back to $\mathbf{T}$ we obtain a prime ideal $\mathscr{P}$ of T.

The problem stated above now amounts to determining whether or not $\mathscr{P}$ and $\mathscr{P}_X$ are primes of fusion. We will write η for the operator $T_M^2 - 1$, viewed either as an element of $\mathbf{T}$ or an element of $\mathbf{T}_X$. Then we wish to study

PROBLEM 2.3. Suppose that $\mathscr{P}_X$ is a prime ideal of $\mathbf{T}_X$ which contains η. Is $\mathscr{P}_X$ necessarily a prime of fusion?

It is easy to see that the answer to this problem is, in general, negative. For example, taking $N = 11$ and $M = 7$, one can prove that the ideal $(T_M + 1)$ of $\mathbf{T}_X$ is a prime ideal containing $T_M^2 - 1$ which is *not* a prime of fusion. In this case, the residue field of our ideal is the field with 5 elements. We will see, however, that Problem 2.3 has an affirmative solution if we impose on $\mathscr{P}_X$ a suitable additional condition, for example that $\mathscr{P}_X$ be prime to the order of the *Shimura subgroup* $\mathbf{S}$ of $J_0(N)$ (as defined in § 4 below). If N is a prime number, this group has order $n = \operatorname{num}\left(\frac{N-1}{12}\right)$ (cf. [7], Ch. II, § 11); if $N = 11$, this group has order 5. Notice especially that the Shimura subgroup of $J_0(N)$ depends only on N, and not on the prime number M.

3. Our result arises from the study of a certain $\mathbf{T}_X$-module Ω, which is closely related to the group $\tilde{L}/L$ mentioned above. Its support contains only primes of fusion, and in fact contains all such primes which do not divide the integer NM. We shall exhibit a $\mathbf{T}_X$-module Δ, whose support consists precisely of the primes containing η, which is furnished with a filtration

$$\Delta = M_0 \supset M_1 \supset M_2 \supset M_3 = 0$$

such that M_1/M_2 is isomorphic to Ω and such that the quotients M_0/M_1 and M_2/M_3 have the same cardinality as $\mathbf{S}$. Any prime ideal of $\mathbf{T}_X$ which contains η and which is prime to the cardinality of $\mathbf{S}$ is consequently in the support of Ω and is therefore a prime of fusion.

Recall ([8], § 2a) the two natural "degeneracy" maps

$$B_1, B_M\colon X_0(NM) \to X_0(N)$$

which correspond, respectively to the identity map and the map $\tau \mapsto M\tau$ on the Poincaré upper half plane. By Pic functoriality they induce maps

$$B_1^*, B_M^*: J_0(N) \to J_0(NM)$$

from which we obtain a homomorphism

$$\alpha: J_0(N) \times J_0(N) \to J_0(NM)$$

by adding these two maps together. The kernel of α is a ceratin finite subgroup Σ of $J_0(N)^2$, which we will study below. The image of α is an abelian subvariety A of $J_0(NM)$ which naturally corresponds to the subspace X of S. For example, when we view $\mathbf{T}$ as a subring of End $(J_0(NM))$ in the usual way we find that $\mathbf{T}$ preserves A and that its action on A factors through $\mathbf{T}_X$.

Similarly, we consider the transpose

$$\alpha': J_0(NM) \to J_0(N) \times J_0(N)$$

of α; it corresponds to the two degeneracy maps induced by B_1 and B_M using Albanese functoriality of the Jacobian. Its kernel is not necessarily connected; in fact, it is an extension of a finite group, canonically isomorphic to the Cartier dual of Σ, by an abelian subvariety B of $J_0(NM)$. The variety B analogously corresponds to Y, so that the action of $\mathbf{T}$ on B factors through $\mathbf{T}_Y$. The intersection

$$\Omega = A \cap B$$

is a finite subgroup of $J_0(NM)$, stable under $\mathbf{T}$, such that the action of $\mathbf{T}$ on Ω factors through both rings $\mathbf{T}_X$ and $\mathbf{T}_Y$. Therefore any prime in the support of Ω is a prime of fusion; and as mentioned above one can show that all primes of fusion occur in the support of Ω, with the possible exceptions of those whose residue fields are of characteristic dividing NM.

Now let L be the line bundle on A arising from the canonical "theta divisor" on $J_0(NM)$ and the inclusion ι of A in $J_0(NM)$. Then L induces an isogeny

$$\phi_L: A \to A^\vee,$$

where $A^\vee$ denotes the abelian variety dual to A. We will denote by $K(L)$ the kernel of this map. It is easy to check the equality

$$\Omega = K(L). \tag{3.1}$$

Indeed, B is quickly seen to be the kernel of the composite $\iota^\vee \circ \phi$, where ϕ is the canonical autoduality of the Jacobian $J_0(NM)$. This gives that Ω

is the kernel of

$$\iota^{\vee}\circ\phi\circ\iota,$$

which is just another way of writing ϕ_L. From (3.1) we obtain a slightly different way of viewing Ω, as follows. Let β denote the isogeny

$$J_0(N)\times J_0(N)\to A$$

for which $\alpha=\iota\circ\beta$. Pulling L back to $J_0(N)^2$ via β, we obtain a line bundle β^*L, whence a finite subgroup $K(\beta^*L)$ of $J_0(N)^2$. This subgroup, which we call Δ, contains Σ and is endowed with a canonical non-degenerate alternating $\mathbf{G}_m$-valued pairing. Let $\Sigma^{\perp}$ be the orthogonal to Σ relative to this pairing; this subgroup of Δ contains Σ, and we have the formula

$$\Omega=\Sigma^{\perp}/\Sigma, \tag{3.2}$$

in view of [9], § 23, Lemma 2.

On the other hand, we can check that Δ is just the kernel of $\alpha'\circ\alpha$. Viewing this endomorphism of $J_0(N)^2$ as a 2×2 matrix of endomorphisms of $J_0(N)$, we find the formula

$$\alpha'\circ\alpha=\begin{bmatrix}M+1 & \tau\\ \tau & M+1\end{bmatrix},$$

where τ is the usual Hecke operator T_M on $J_0(N)^2$. In other words, we have

$$\Delta=\{(x,y)\in J_0(N)^2 \mid \tau x=-(M+1)y \text{ and } \tau y=-(M+1)x\}.$$

We now claim that $\mathbf{T}_X$ acts on $J_0(N)^2$ as a subring of endomorphisms of this abelian variety. By this we mean that, for each each $n\geqslant 1$, the quantity

$$T'_n=\beta^{-1}\circ T_n\circ\beta,$$

a priori an endomorphism of $J_0(N)^2$ up to isogeny, is in fact a genuine endomorphism of $J_0(N)^2$. This assertion is clear indeed if n is prime to M; in that case, T'_n is nothing but the usual Hecke operator T_n on $J_0(N)$, acting "diagonally" on the product $J_0(N)^2$. Thus the general case follows from the explicit formula

$$T'_M=\begin{bmatrix}\tau & M\\ -1 & 0\end{bmatrix}.$$

In what follows, we will omit the superscript $'$ and write T_n for T'_n.

PROPOSITION 3.3. *The group Δ is the kernel of the endomorphism $\eta=T_M^2-1$ of $J_0(N)^2$.*

This proposition is proved by a short calculation, which we omit; the reader can verify that we have more precisely the identity

$$\eta = \begin{bmatrix} -1 & \tau \\ 0 & -1 \end{bmatrix} \circ (\alpha' \circ \alpha).$$

Because of (3.3), we can view Δ as a $\mathbf{T}_X$-module. Its *support* consists of all prime ideals of $\mathbf{T}_X$ which contain the annihilator I of Δ in $\mathbf{T}_X$. We have

$$I = \{T \in \mathbf{T}_X |\ T = \varepsilon\eta \text{ for some } \varepsilon \in \mathrm{End}(J_0(N)^2)\},$$

i.e.,

$$I = R\eta \cap \mathbf{T}_X,$$

where

$$R = (\mathbf{T}_X \otimes \mathbf{Q}) \cap \mathrm{End}(J_0(N)^2).$$

Now R is a finitely generated abelian group, so certainly finitely generated as a $\mathbf{T}_X$-module. Hence if $\mathscr{P}_X$ is a maximal ideal of $\mathbf{T}_X$ we have

$$\mathscr{P}_X = R\mathscr{P}_X \cap \mathbf{T}_X.$$

This formula shows that we have $\mathscr{P}_X \supseteq I$ if and only if $\mathscr{P}_X$ contains η. Hence we get:

PROPOSITION 3.4. *The support of Δ consists precisely of those primes of $\mathbf{T}_X$ which contain η.*

This proposition may be viewed as a partial solution to Problem 2.3. It fails to be a complete solution because of the group Σ, which is the obstruction to the equality between Δ and Ω. We will determine Σ in the next section.

4. Our analysis of Σ is based on results contained in Ihara's article [6].[1] We will find, in studying Σ, that the analogue of this group is 0 in the situation where $\Gamma_0(N)$ is replaced by either of its subgroups $\Gamma_1(N)$ or $\Gamma(N)$. For definiteness in what follows, we shall regard these groups as subgroups of $\mathbf{PSL}(2, \mathbf{Z})$, rather than $\mathbf{SL}(2, \mathbf{Z})$. We will let $X_1(N)$ and $X(N)$ be the modular curves associated with these groups and let $J_1(N)$ and $J(N)$ be, as usual, the Jacobians of these curves.

[1] The author wishes to thank J.-P. Serre for bringing Ihara's results to his attention.

The inclusions $\Gamma(N) \subset \Gamma_1(N)$ and $\Gamma_1(N) \subset \Gamma_0(N)$ correspond to coverings of curves

$$\sigma\colon X(N) \to X_1(N), \qquad \pi\colon X_1(N) \to X_0(N).$$

From these maps we obtain, by Pic functoriality, maps

$$\pi^*\colon J_0(N) \to J_1(N), \qquad \sigma^*\colon J_1(N) \to J(N).$$

The kernel of π^* is a finite subgroup S of $J_0(N)$ which is known as the *Shimura subgroup* of $J_0(N)$. It is isomorphic to the $\mathbf{G}_m$-dual of the covering group of the maximal abelian unramified covering of $X_0(N)$ which is intermediate to $X_1(N) \to X_0(N)$. The kernel of σ^* is trivial, because there are no unramified coverings of $X_1(N)$ intermediate to $X(N) \to X_1(N)$: the standard cusp "∞" of $X_1(N)$ is totally ramified in this covering.

If Γ is one of the three groups $\Gamma(N)$, $\Gamma_1(N)$, $\Gamma_0(N)$, we will let Γ' be the intersection of Γ with $\Gamma_0(M)$. We use the symbol $'$ in writing the corresponding modular curves and their Jacobians. Thus $J_0'(N)$, for instance, is $J_0(NM)$. We then have a pair of commutative diagrams

$$\begin{array}{ccc} J(N)^2 & \xrightarrow{\gamma} & J'(N) \\ {\scriptstyle\sigma^*}\uparrow & & \uparrow \\ J_1(N) & \xrightarrow{\beta} & J_1'(N) \end{array} \qquad \begin{array}{ccc} J_1(N)^2 & \xrightarrow{\beta} & J_1'(N) \\ {\scriptstyle\pi^*}\uparrow & & \uparrow \\ J_0(N)^2 & \xrightarrow{\alpha} & J_0'(N) \end{array}$$

in which the horizontal maps are the obvious degeneracy maps. We shall admit for the moment the following result:

THEOREM 4.1. *The kernel of γ is trivial.*

Then, by the above discussion, we certainly have:

COROLLARY 4.2. *The kernel of β is trivial. The kernel Σ of α is a subgroup of* $\mathrm{S} \times \mathrm{S}$.

More precisely, we will prove

THEOREM 4.3. *The group Σ is the subgroup*

$$T = \{(x, y) \in \mathrm{S} \times \mathrm{S} \mid x + y = 0\}$$

of $\mathrm{S} \times \mathrm{S}$.

To prove Theorem 4.3 we first will show that Σ contains T. Let

$$B\colon J_0(N) \to J_0(NM)$$

be the degeneracy map B_1^*. Then the degeneracy map B_M^* is the composition $W_M \circ B$, where W_M is the indicated Atkin–Lehner involution of $J_0(NM)$. The inclusion $\Sigma \supseteq T$ thus means that W_M acts on the group $B(\mathbf{S})$ by multiplication by $+1$. Now the Atkin–Lehner involution W_N on $J_0(N)$ acts on $\mathbf{S}$ by multiplication by -1 (cf. [7], Chapter II, Proposition 11.7), which gives that the Atkin–Lehner involution W_N of $J_0(NM)$ acts on $B(\mathbf{S})$ by multiplication by -1. But since $B(\mathbf{S})$ is a subgroup of the Shimura subgroup of $J_0(NM)$, we find that W_{NM} acts on $B(\mathbf{S})$ by multiplication by -1. Since $W_M = W_{NM} \circ W_N$, we get that Σ contains T.

In view of this inclusion, the assertion $\Sigma = T$ amounts to the injectivity of B on $\mathbf{S}$. In fact, B has kernel 0 because the covering $B_1\colon X_0(NM) \to X_0(N)$ is ramified and such that there is no non-trivial covering of $X_0(N)$, other than B_1, which is intermediate to B_1.

Proof of Theorem 4.1. We must show, for all prime numbers l, the injectivity of the map

$$H^1(X(N), \mathbf{Z}/l\mathbf{Z}) \oplus H^1(X(N), \mathbf{Z}/l\mathbf{Z}) \to H^1(X'(N), \mathbf{Z}/l\mathbf{Z})$$

resulting from the two degeneracy coverings $X'(N) \rightrightarrows X(N)$. We may view $H^1(X(N), \mathbf{Z}/l\mathbf{Z})$ as classifying unramified Galois coverings of $X(N)$ with structural group $\mathbf{Z}/l\mathbf{Z}$, and the problem is to show that there is no non-trivial pair of such coverings which become equal after pullback to $X'(N)$ by the two different degeneracy maps. However, this is just a special case of Lemma 3.2 of Ihara [6], which asserts that the system

$$\begin{array}{ccc} & X'(N) & \\ \swarrow & & \searrow \\ X(N) & & X(N) \end{array}$$

is "simply connected". (See also the remarks at the beginning of § 3.4 of [6].)

Alternatively, we may obtain a slightly more direct proof of Theorem 4.1 from the ingredients used in the proof of Ihara's Lemma 3.2. Here we view Theorem 4.1 as asserting the surjectivity of the natural map

$$H_1(X'(N), \mathbf{Z}) \to H_1(X(N), \mathbf{Z}) \oplus H_1'(X(N), \mathbf{Z}).$$

In terms of subgroups of $\mathbf{PSL}(2, \mathbf{Q})$, we have corresponding inclusions of the group

$$\Delta = \Gamma(N) \cap \Gamma_0(N)$$

in the two groups

$$G_1 = \Gamma(N), \quad G_2 = \begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix}^{-1} \Gamma(N) \begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix};$$

we wish to prove the surjectivity of

$$H_1(A, \mathbf{Z}) \to H_1(G_1, \mathbf{Z})/\mathrm{Pb}(G_1) \oplus H_1(G_2, \mathbf{Z})/\mathrm{Pb}(G_2). \tag{4.4}$$

(Here $\mathrm{Pb}(G_i)$ denotes the subgroup of $H_1(G_i, \mathbf{Z})$ generated by the set of parabolic elements of the group G_i, for $i = 1, 2$.)

Let Γ be the principal congruence subgroup of level N (i.e., the analogue of $\Gamma(N)$) in $\mathbf{PSL}(2, \mathbf{Z}[1/M])$. Then the inclusions of G_1 and G_2 in Γ are well known to induce an isomorphism of the amalgamated product $G_1 *_A G_2$ with Γ (see, e.g., [11], Ch. II, § 1.4). Accordingly, by the exact sequence of Lyndon (see, e.g., [*loc. cit.*], page 169), the cokernel of the map

$$H_1(A, \mathbf{Z}) \to H_1(G_1, \mathbf{Z}) \oplus H_1(G_2, \mathbf{Z})$$

may be identified with $H_1(\Gamma, \mathbf{Z})$. Let Δ be the subgroup of Γ generated by the commutator subgroup of Γ and by the parabolic elements of G_1 and G_2. Then the cokernel of the map (4.4) may be identified with the quotient Γ/Δ.

Since Γ is generated by its parabolic elements, the surjectivity of (4.4) thus means that all parabolic elements of Γ lie in Δ. As on page 178 of [6], we now note that if γ is a parabolic element of Γ, then γ^{M^n} lies in G_1 for some positive integer n. It is easy to deduce from this that γ lies in Δ (*loc. cit.*).

References

[1] Doi K. and Hida H., *On a Certain Congruence of Cusp Forms and the Special Values of Their Dirichlet Series*, Unpublished manuscript, 1979.

[2] Doi K. and Ohta M., *On Some Congruences Between Cusp Forms on $\Gamma_0(N)$*, Lecture Notes in Math. 601, pp. 91–105, Springer, Berlin–Heidelberg–New York, 1977.

[3] Hida H., Congruences for Cusp Forms and Special Values of Their Zeta Functions, *Invent. Math.* **63** (1981), pp. 225–261.

[4] Hida H., On Congurence Divisors of Cusp Forms as Factors of the Special Values of Their Zeta Functions, *Invent. Math.* **64** (1981), pp. 221–262.

[5] Hida H., Kummer's Criterion for the Special Values of Hecke L-Functions of Imaginary Quadtratic Fields and Congruences Among Cusp Forms, *Invent. Math.* **66** (1982), pp. 415–459.

[6] Ihara Y., On Modular Curves Over Finite Fields, *Proceedings of the International Colloquium on Discrete Subgroups of Lie Groups and Applications to Moduli*, Bombay, January 1973, pp. 161–202.

[7] Mazur B., Modular Curves and the Eisenstein Ideal, *Publ. Math. IHES* **47** (1977), pp. 33–186.
[8] Mazur B., Rational Isogenies of Prime Degree, *Invent. Math.* **44** (1978), pp. 129–162.
[9] Mumford D., *Abelian Varieties*, Oxford University Press, London, 1970.
[10] Ribet K., Mod p Hecke Operators and Congruences Between Modular Forms, *Invent. Math.* **71** (1983), pp. 193–205.
[11] Serre J.-P., Arbres, Amalgames, SL_2, *Astérisque* **46**, Société Mathématique de France, Paris, 1977.

U.C. BERKELEY MATH. DEPARTMENT,
BERKELEY, CA.
94720 USA

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

WOLFGANG M. SCHMIDT

Analytic Methods for Congruences, Diophantine Equations and Approximations

Our present methods for estimating the number of solutions of congruences, equations and inequalities have a number of common features. In particular, these methods involve exponential sums.

1. Congruences and exponential sums

Consider a system of congruences

$$\mathfrak{F}_i(x_1, \ldots, x_s) \equiv 0 \pmod m \qquad (i = 1, \ldots, r) \tag{1.1}$$

where each $\mathfrak{F}_i$ is a polynomial with integer coefficients. We are interested in the number $N = N(\mathfrak{F}_1, \ldots, \mathfrak{F}_r) = N(\underline{\mathfrak{F}})$ of solutions, and more generally in the number $N_{\mathfrak{D}} = N_{\mathfrak{D}}(\underline{\mathfrak{F}})$ of solutions $\boldsymbol{x} = (x_1, \ldots, x_s)$ lying in a given subset $\mathfrak{D}$ of $(\boldsymbol{Z}/m\boldsymbol{Z})^s$. It is well known that

$$N_{\mathfrak{D}} = m^{-r} \sum_{\boldsymbol{a} (\operatorname{mod} m)} E_{\mathfrak{D}}(\boldsymbol{a}\underline{\mathfrak{F}}), \tag{1.2}$$

where $\boldsymbol{a} = (a_1, \ldots, a_r)$, $\boldsymbol{a}\underline{\mathfrak{F}} = a_1\mathfrak{F}_1 + \ldots + a_r\mathfrak{F}_r$, and

$$E_{\mathfrak{D}}(\mathfrak{F}) = \sum_{\boldsymbol{x} \in \mathfrak{D}} e\big(m^{-1}\mathfrak{F}(\boldsymbol{x})\big) \tag{1.3}$$

with $e(z) = e^{2\pi i z}$. Progress has recently been made in estimating $E_{\mathfrak{D}}$ and $N_{\mathfrak{D}}$.

First of all, the fundamental work of Deligne based on algebraic geometry continues to have applications. Suppose that $m = p$, a prime, and that $\mathfrak{F}$ is of degree d with $p \nmid d$, and such that the homogeneous part of $\mathfrak{F}$ of degree d, call it $\mathfrak{F}^{(d)}$, is nonsingular. It had already been shown in 1974 [10] that under these conditions

$$|E(\mathfrak{F})| \leqslant (d-1)^s p^{s/2}, \tag{1.4}$$

where E is the "complete sum", i.e., the sum $E_{\mathfrak{D}}$ with $\mathfrak{D} = (\boldsymbol{Z}/p\boldsymbol{Z})^s$. Many other important and general results were proved by Deligne and Katz. I may refer, e.g., to the exposition [12] of Katz.

Very recently [24] I could prove novel results by "elementary" methods. Still let $m = p$, a prime. Let $\mathfrak{F}$ be a polynomial over the field $\boldsymbol{F}_p = \boldsymbol{Z}/p\boldsymbol{Z}$ of degree $d > 1$ with homogeneous part $\mathfrak{F}^{(d)}$. Write $h = h(\mathfrak{F}) = h(\mathfrak{F}^{(d)})$ for the least integer such that $\mathfrak{F}^{(d)}$ may be written as

$$\mathfrak{F}^{(d)} = \mathfrak{A}_1\mathfrak{B}_1 + \ldots + \mathfrak{A}_h\mathfrak{B}_h, \tag{1.5}$$

where the $\mathfrak{A}_i, \mathfrak{B}_i$ are forms of positive degrees with coefficients in $\boldsymbol{F}_p$. Then

$$|E(\mathfrak{F})| \ll p^{s-\varkappa} \tag{1.6}$$

where $\varkappa = h/\psi(d)$ with $\psi(2) = 2$, $\psi(3) = 4$, $\psi(4) = 24$, $\psi(5) = 208$, and with $\psi(d) < 3^d \cdot d!$ in general. The constant in $\ll$ here depends only on s and d. The first nontrivial case, namely $d = 3$, is due to Davenport and Lewis [9]. I expect that eventually (1.6) will be improved quantitatively, by replacing $\psi(d)$ by a smaller function, and perhaps qualitatively, by replacing h by other invariants.

Suppose now that $\mathfrak{D}$ is a box $\mathfrak{B} = \mathfrak{J}_1 \times \ldots \times \mathfrak{J}_s$, where each $\mathfrak{J}_j$ is an interval, i.e., the set of residue classes of integers in $a_j < x_j \leqslant b_i$, say. We will say that $\mathfrak{B}$ is a box of size $\leqslant P$ if $b_j - a_j \leqslant P$ $(j = 1, \ldots, s)$, and a box of size $\geqslant P$ if $b_j - a_j \geqslant P$ $(j = 1, \ldots, s)$. As was pointed out by Serre [25], the estimates of Deligne on complete sums give nontrivial results for incomplete sums over boxes of size $\leqslant P$, provided P is somewhat larger than $p^{1/2}$. When $d > 2$, the elementary methods allow much smaller boxes. If $\mathfrak{F}$ and h are as above, let $\mathfrak{B}$ be a box of size $\leqslant P$ where $P = p^\delta$ with $d^{-1} < \delta \leqslant 1$. Then

$$|E_{\mathfrak{B}}(\mathfrak{F})| \ll P^{s-\varkappa(\delta)+\varepsilon}, \tag{1.7}$$

where

$$\varkappa(\delta) = (d-\delta^{-1})(d-1)^{-1}\varkappa = (d-\delta^{-1})(d-1)^{-1}\psi(d)^{-1}h,$$

where $\varepsilon > 0$, and where the constant in $\ll$ depends only on s, d, δ, ε.

2. Small solutions of congruences

The above results may be used to show that certain rather small boxes contain solutions of systems of congruences. Using (1.4), Myerson [16] showed that a system (1.1) with $m = p$ has a solution in every box $\mathfrak{B}$ of cardinality

$$|\mathfrak{B}| \geqslant \big((1+\varepsilon)\,(2d-2)\big)^s p^{(s/2)+r}$$

for $\varepsilon > 0$ and $p > c_1(s, d, r, \varepsilon)$, provided that: (A) the polynomials of the pencil of $\underline{\mathfrak{F}} = (\mathfrak{F}_1, \ldots, \mathfrak{F}_r)$, i.e., the polynomials $\boldsymbol{a}\underline{\mathfrak{F}}$ with $\boldsymbol{a} \in \boldsymbol{F}_p^r \setminus \boldsymbol{0}$, have degrees between 2 and d, (B) the homogeneous part of highest degree of each $\boldsymbol{a}\underline{\mathfrak{F}}$ is nonsingular, and (C) $\mathfrak{F}_1 = \ldots = \mathfrak{F}_r = 0$ defines an absolute variety of dimension $s-r$. In particular, under these assumptions, there is a solution with

$$0 \leqslant x_j \leqslant 2(1+\varepsilon)p^{(1/2)+(r/s)} \qquad (j = 1, \ldots, s). \tag{2.1}$$

The exponent here may be improved, at least in certain cases. R. C. Baker [4] has shown for a single congruence $\mathfrak{F}(\boldsymbol{x}) \equiv 0 \pmod m$ that there is a solution $\boldsymbol{x} \neq \boldsymbol{0}$ with

$$|x_j| \leqslant m^{(1/2)+(1/2s-2)+\varepsilon} \qquad (j = 1, \ldots, s), \tag{2.2}$$

provided that the reduction of $\mathfrak{F}$ modulo each prime factor p of m is a nonsingular form of degree $d > 1$, and that $\varepsilon > 0$ and $m > c_2(s, d, \varepsilon)$. He further has shown that for $m > c_3(s, d, \varepsilon)$, a diagonal congruence

$$a_1 x_1^d + \ldots + a_s x_s^d \equiv 0 \pmod m$$

has a nontrivial solution $\boldsymbol{x}$ with (2.2).

Further applications of the results from algebraic geometry were made by Hooley. I may refer to his contribution to the present volume.

On the other hand, the elementary estimate (1.7) has the following consequences. Let the modulus m again be a prime p. Let $\underline{\mathfrak{F}} = (\mathfrak{F}_1, \ldots, \mathfrak{F}_r)$ be an r-tuple of polynomials with coefficients in $\boldsymbol{F}_p$ such that every polynomial in the pencil has a degree between 2 and d. Let

$$h = h(\underline{\mathfrak{F}})$$

signify the minimum of $h(\mathfrak{F})$ over the polynomials $\mathfrak{F}$ of the pencil. Now if $\mathfrak{B}$ is a box of size $\leqslant P = p^\delta$ with $d^{-1} < \delta \leqslant 1$, then

$$N_{\mathfrak{B}}(\underline{\mathfrak{F}}) = p^{-r}|\mathfrak{B}| + O(P^{s-\varkappa+\varepsilon}) \tag{2.3}$$

for $\varepsilon > 0$, with a constant in O which depends only on $s, d, r, \delta, \varepsilon$. In particular, every box $\mathfrak{B}$ of size $\geqslant p^\delta$ contains a solution, provided that

$$h > r(d-1)\,(d\delta-1)^{-1}\psi(d), \tag{2.4}$$

and that $p > c_4(s, d, r, \delta, \varepsilon)$.

It may be deduced that when $\mathfrak{F}_1, \ldots, \mathfrak{F}_r$ are arbitrary forms of degree $\leqslant d$, then the congruences (1.1) with $m = p$ have a solution $\boldsymbol{x} \neq \mathbf{0}$ with

$$|x_j| \leqslant p^{(1/2)+\varepsilon} \quad (j = 1, \ldots, s), \tag{2.5}$$

provided $\varepsilon > 0$ and $s > c_5(d, r, \varepsilon)$ and $p > c_6(d, r, \varepsilon)$. The 1/2 in the exponent here is clearly best possible for forms of even degree, but it (the 1/2) may be removed altogether (see § 4 below) when all the forms are of odd degree.

Very likely, a result like (2.5) is true for arbitrary modulus m. This is in fact fairly easy to see for a single quadratic form (Schinzel, Schlickewei and Schmidt [17]), and was established for systems of quadratic forms by R. C. Baker [3]. Baker proved a weaker result for a quartic form and arbitrary modulus.

Added in proof: (2.5) was extended to an arbitrary modulus by the author: "Small solutions of congruences in a large number of variables" (submitted). (2.5) with prime modulus holds for a quadratic form and $s \geqslant 4$, as shown by D. R. Heath-Brown.

3. p-adic equations

Suppose now that $m = p^l$ where p is a fixed prime. As is well known, the solubility of our system (1.1) for $l = 1, 2, \ldots$ is equivalent to the solubility of the system of equations

$$\mathfrak{F}_i(x_1, \ldots, x_s) = 0 \quad (i = 1, \ldots, r) \tag{3.1}$$

in p-adic integers $x_1, \ldots, x_s$. Moreover, when each $\mathfrak{F}_i$ is a form, the existence of a *primitive* solution of (1.1), i.e., a solution $\boldsymbol{x}$ with $\boldsymbol{x} \not\equiv 0 \pmod{p}$, for $l = 1, 2, \ldots$, is equivalent to the nontrivial solubility of (3.1) in p-adic numbers.

E. Artin had conjectured that r forms of degree d in more than $d^2 r$ variables always have a nontrivial p-adic zero. Put differently, the conjecture was that $v_{d,r} \leqslant d^2 r$, where $v_{d,r}$ is the least number such that any r forms of degree d in more than $v_{d,r}$ variables possess a nontrivial p-adic zero for each prime p. Since it may be seen that $v_{d,r} \geqslant d^2 r$, the conjecture could be restated as $v_{d,r} = d^2 r$. It is true for $v_{2,1}$, $v_{2,2}$, $v_{3,1}$. But Terjanian [26] disproved it for $v_{4,1}$, and more recently [27] he even showed that $v_{4,1} \geqslant 20$. For a while some people thought that perhaps $v_{d,1}$ may grow not faster than d^3. But Arhipov and Karačuba [1] demonstrated that for $\varepsilon > 0$ and some arbitrarily large values of d we have

$$v_{d,1} > e^{d/(\log d)^{2+\varepsilon}}.$$

This was improved independently by Arhipov and Karačuba themselves

[2], as well as by Brownawell [6], and by Lewis and Montgomery [15], to

$$v_{d,1} > e^{d/(\log d)^{1+\varepsilon}}. \tag{3.2}$$

So far, no counterexample to Artin's conjecture has been found with odd d. Hence for all that we know, the conjecture may still be true for odd d. As for upper bounds, the best that is known seems to be my estimate [22]

$$v_{d,1} < e^{2^{d}\cdot d!}, \tag{3.3}$$

valid for $d > d_0$.

It is perhaps a little easier to keep d fixed and to bound $v_{d,r}$ as a function of r. Leep [13] has shown that $v_{2,r} \leqslant 2r^2+2r-4$ when $r \geqslant 2$. Extending Leep's elementary diagonalization method, Leep and I [14] established that $v_{3,r} < (81/2)r^4$, and in general that

$$v_{d,r} < c_7(d)r^{2^{d-1}}. \tag{3.4}$$

In the case where $d = 3$, I used heavy machinery to establish [20] the stronger bound

$$v_{3,r} \ll r^3.$$

Now let z_l be the number of solutions of (1.1) with $m = p^l$. The limit

$$\lim_{l\to\infty} z_l/p^{l(s-r)},$$

when it exists, is called the *p-adic density* of zeros of $\underline{\mathfrak{F}}$. Conditions for the existence of this limit and estimates for this limit were given in [19] when $\underline{\mathfrak{F}}$ is a system of quadratic forms, and in [20] when $\underline{\mathfrak{F}}$ is a system of cubic forms.

4. Diophantine equations

Consider now a system of equations

$$\mathfrak{F}_i(x_1, \ldots, x_s) = 0 \quad (i = 1, \ldots, r) \tag{4.1}$$

where each $\mathfrak{F}_i$ is a polynomial with integer coefficients. Given a finite set $\mathfrak{D}$ of integer points, i.e., a finite subset of $\boldsymbol{Z}^s$, write $M_{\mathfrak{D}} = M_{\mathfrak{D}}(\underline{\mathfrak{F}})$ for the number of solutions of (4.1) with $\boldsymbol{x} \in \mathfrak{D}$. We have

$$M_{\mathfrak{D}} = \int_U S_{\mathfrak{D}}(\boldsymbol{a}\underline{\mathfrak{F}})\,d\boldsymbol{a}, \tag{4.2}$$

where $\boldsymbol{a} = (a_1, \ldots, a_r)$, $\boldsymbol{a}\underline{\mathfrak{F}} = a_1\mathfrak{F}_1 + \ldots + a_r\mathfrak{F}_r$, the integration is over the unit cube U: $0 \leqslant a_i < 1$ $(i = 1, \ldots, r)$, and $S_{\mathfrak{D}}$ is the exponential sum

$$S_{\mathfrak{D}}(\mathfrak{F}) = \sum_{\boldsymbol{x}\in\mathfrak{D}} e(\mathfrak{F}(\boldsymbol{x})). \tag{4.3}$$

The well-known formula (4.2) corresponds to (1.2), and it is the basis of the Hardy–Littlewood "Circle Method".

In classical work the formula was applied in cases where the polynomials $\mathfrak{F}_i$ were diagonal forms. In 1957 Birch [5] used a diagonalization process to show that a system of r homogeneous diophantine equations of odd degree d has a nontrivial solution if the number s of variables exceeds a certain number $w_{d,r}$. More recently [18] I gave the following refinement: When $\varepsilon > 0$ and the number of variables exceeds $w_{d,r}(\varepsilon)$, then there is a nontrivial solution with $|x_j| \leqslant F^{\varepsilon}$ $(j = 1, \ldots, s)$, where F is the maximum modulus of the coefficients of the forms $\mathfrak{F}_i$.

In Birch's result it is of course necessary to suppose that d is odd. In the case of an indefinite quadratic form, i.e., a form with a nontrivial *real* zero, we know by Meyer's Theorem that a nontrivial integer zero does exist if the number of variables is at least 5. No corresponding result holds for quartic forms. For the form

$$(x_1^2 + x_2^2 + \ldots + x_{s-1}^2)^2 - 2x_s^4$$

has no nontrivial integer zero, no matter how large s is, even though it has a nontrivial real zero, and, when $s \geqslant 6$, a nontrivial p-adic zero for each prime p.

Now let $w_{d,r}$ be the smallest number for which the assertion in Birch's Theorem holds. Since a form $\mathfrak{F}$ of odd degree $l < d$ may be replaced by the form $(x_1^2 + \ldots + x_s^2)^{(d-l)/2}\mathfrak{F}$ of degree d, this number $w_{d,r}$ is smallest possible such that every system of r forms of odd degrees *at most* d in more than $w_{d,r}$ variables has a nontrivial zero. It is possible in principle to derive upper bounds for $w_{d,r}$ by Birch's method. But the bounds so obtained would be extremely poor, and it remains a task of considerable interest to find reasonable bounds.

In a series of papers beginning in 1959, Davenport applied (4.2) directly to a general cubic form. In [7] he showed that $w_{3,1} \leqslant 15$. It is known that $w_{3,1} \geqslant 9$, and it is generally believed that $w_{3,1} = 9$. Heath-Brown [11] in deep work which uses Deligne's estimates of exponential sums, and which has a number of novel features, proved that a *nonsingular* cubic form in more than 9 variables possesses a nontrivial zero. On the other hand, I proved [21] that $w_{3,r} \ll r^5$, more precisely that $w_{3,2} \leqslant 5139$ and $w_{3,r} < (10r)^5$. I conjecture that in analogy to (3.4) we have

$$w_{d,r} < c_8(d)\, r^{c_9(d)}$$

for each odd d. But for $d = 5$ we do not even know whether $w_{5,r} < e^{c_{10}r}$; we only know [23] that $\log\log w_{5,r} \ll r$.

We end this section with comments about the "density" of solutions. Put

$$M_P = M_P(\underline{\mathfrak{F}}) = M_{\mathfrak{D}(P)}(\underline{\mathfrak{F}}),$$

where $\mathfrak{D}(P)$ is the cube of points x with $|x_j| \leqslant P$ $(j = 1, \ldots, s)$. Under suitable conditions on $\underline{\mathfrak{F}}$ one may derive an asymptotic formula for M_P.

For simplicity suppose that $r = 1$ and that $\mathfrak{F}$ is a form of degree $d > 1$. Define $h = h(\mathfrak{F})$ much as in § 1, i.e., as the least number such that $\mathfrak{F}$ may be written as $\mathfrak{F} = \mathfrak{A}_1\mathfrak{B}_1 + \ldots + \mathfrak{A}_h\mathfrak{B}_h$ with forms $\mathfrak{A}_i$, $\mathfrak{B}_i$ of positive degrees with *rational* coefficients. Now if

$$h \geqslant \omega(d)$$

where $\omega(2) = 4$, $\omega(3) = 16$, $\omega(4) = 144$, and in general $\omega(d) \leqslant 3^d \cdot d!$, then [23]

$$M_P(\mathfrak{F}) = \mu P^{s-d} + O(P^{s-d-\delta}) \tag{4.4}$$

where $\delta > 0$ and μ depend only on $\mathfrak{F}$. Here μ is the product of p-adic densities as defined in § 3 and a suitable real density. Moreover, $\mu > 0$ if $h \geqslant \omega^*(d)$ for a certain function ω^* and if the manifold of real zeros of $\mathfrak{F}$ has dimension $s-1$. This last condition is always fulfilled when d is odd.

By using a generalized version of this result, one can show that there is a constant $w^* = w^*_{d,r}$ such that every system $\underline{\mathfrak{F}}$ of r forms of odd degrees at most d has

$$M_P(\underline{\mathfrak{F}}) \gg P^{s-w^*} \tag{4.5}$$

when $s > w^*$. The constant in $\gg$ here depends only on $\underline{\mathfrak{F}}$.

This is another refinement of the Theorem of Birch. An analysis of Davenport's method gives $M_P(\mathfrak{F}) \gg P^{s-15}$ for a cubic form $\mathfrak{F}$. We have $w_{d,r} \leqslant w^*_{d,r}$, and at present our knowledge of $w^*_{d,r}$ is as good (i.e., as bad) as our knowledge of $w_{d,r}$.

5. Diophantine inequalities

We now turn to systems of inequalities

$$|\mathfrak{F}_i(x)| < \varepsilon \quad (i = 1, \ldots, r), \tag{5.1}$$

where each $\mathfrak{F}_i$ is a polynomial with *real* coefficients. Given a finite set $\mathfrak{D} \subseteq \boldsymbol{Z}^s$, let $L_{\mathfrak{D}}(\underline{\mathfrak{F}}, \varepsilon)$ be the number of solutions $x \in \mathfrak{D}$ of (5.1). As a matter of fact, it is more convenient to deal with modified functions $L^{(0)}_{\mathfrak{D}}$, $L^{(1)}_{\mathfrak{D}}$. Here $L^{(0)}_{\mathfrak{D}}$ counts not only solutions of (5.1), but also solutions of $|\mathfrak{F}_i(x)| \leqslant \varepsilon$, with weight 2^{-k} if the equality sign holds in k of these relations. On the other hand, $L_{\mathfrak{D}}^{(1)}$ counts the solutions of (5.1) with weight $\prod_{i=1}^{r}\left(1-\varepsilon^{-1}|\mathfrak{F}_i(\boldsymbol{x})|\right)$. Since

$$\int_{-\infty}^{\infty} e(\alpha\theta)\left(\frac{\sin 2\pi\alpha}{\pi\alpha}\right)d\alpha = \begin{cases} 1 & \text{when} \quad |\theta|<1, \\ 0 & \text{when} \quad |\theta|>1, \\ \frac{1}{2} & \text{when} \quad |\theta|=1, \end{cases}$$

it is not difficult to see that

$$L_{\mathfrak{D}}^{(0)} = \int_{\boldsymbol{R}^r} T(\boldsymbol{a})\left(\prod_{i=1}^{r}\frac{\sin 2\pi\alpha_i\varepsilon}{\pi\alpha_i}\right)d\boldsymbol{a}$$

with

$$T(\boldsymbol{a}) = \sum_{\boldsymbol{x}\in\mathfrak{D}} e\left(\boldsymbol{a}\underline{\mathfrak{F}}(\boldsymbol{x})\right).$$

On the other hand, since

$$\int_{-\infty}^{\infty} e(\alpha\theta)\left(\frac{\sin\pi\alpha}{\pi\alpha}\right)^2 d\alpha = \begin{cases} 1-|\theta| & \text{when} \quad |\theta|\leqslant 1, \\ 0 & \text{when} \quad |\theta|\geqslant 1, \end{cases}$$

one has

$$L_{\mathfrak{D}}^{(1)} = \varepsilon^{-r}\int_{\boldsymbol{R}^r} T(\boldsymbol{a})\left(\prod_{i=1}^{r}\frac{\sin\pi\alpha_i\varepsilon}{\pi\alpha_i}\right)^2 d\boldsymbol{a}.$$

These formulae now take the place of (1.2) and (4.2).

It usually is difficult to compute $L_{\mathfrak{D}}^{(0)}$, and there are almost no instances where we know asymptotic formulae for $L_P = L_{\mathfrak{D}(P)}$. It is easier to estimate $L_{\mathfrak{D}}^{(1)}$, and since $L_{\mathfrak{D}} \geqslant L_{\mathfrak{D}}^{(1)}$, this gives lower bounds for $L_{\mathfrak{D}}$. Beginning with work of Davenport and Heilbronn [8] in 1946, a number of results were established by this method, usually involving diagonal forms.

In [18] I proved: Given r forms $\mathfrak{F}_i$ with real coefficients and of odd degree $\leqslant d$, the inequalities (5.1) can be satisfied for each $\varepsilon>0$ by a nonzero integer point $\boldsymbol{x}$, provided only that the number of variables exceeds some number $u_{d,r}$. More precisely, let $u_{d,r}$ be the smallest number with this property. We mentioned above that $w_{3,r} \ll r^5$, but we do not know whether $u_{3,r} \ll r^{c_{11}}$ for some c_{11}, we do not even know whether $u_{3,r} \ll e^r$, or whether $u_{3,r} \ll e^{e^r}$. Nor do we know the analogue of (4.5), i.e., whether there exists an $u^* = u^*_{d,r}$ such that

$$L_P \gg P^{s-u^*}.$$

On the other hand, it is known that when $\mathfrak{F} = (\mathfrak{F}_1, \ldots, \mathfrak{F}_r)$ is a system as before and in $s > u_{d,r}(c)$ variables where $c > 0$, then not only does (5.1) have infinitely many solutions, but there are infinitely many integer points $x \neq 0$ with

$$|\mathfrak{F}_i(x)| < |x|^{-c} \quad (i = 1, \ldots, r), \tag{5.2}$$

where $|x| = \max(|x_1|, \ldots, |x_s|)$.

I close with another conjecture. Write $\|a\|$ for the distance from a real number a to the nearest integer: Given a form $\mathfrak{F}$ of degree d in $s > c_{12}(d, \varepsilon)$ variables, there are infinitely many nonzero integer points x with

$$\|\mathfrak{F}(x)\| < |x|^{-2+\varepsilon}.$$

For odd d this (and much more) is true in view of (5.2). The only other known case is when $d = 2$ [17].

References

[1] Arhipov G. I. and Karačuba A. A., Local representation of zero by a form (in Russian), *Izv. Akad. Nauk SSSR. Ser. Mat.* **45** (1981), pp. 948–961.

[2] Arhipov G. I. and Karačuba A. A., Announcement in *Izv. Akad. Nauk USSR.*

[3] Baker R. C., Small solutions of quadratic and quartic congruences, *Mathematika* **27** (1980), pp. 30–45.

[4] Baker R. C., *Small solutions of congruences* (to appear).

[5] Birch B. J., Homogeneous forms of odd degree in a large number of variables, *Mathematika* **4** (1957), pp. 102–105.

[6] Brownawell D., On p-adic zeros of forms, *J. of Number Theory* (to appear).

[7] Davenport H., Cubic forms in 16 variables, *Proc. Royal Soc.*, **A**, **272** (1963), pp. 285–303.

[8] Davenport H. and Heilbronn H., On indefinite quadratic forms in five variables, *J. London Math. Soc.* **21** (1946), pp. 185–193.

[9] Davenport H. and Lewis D. J., Exponential sums in many variables, *Amer. J. of Math.* **84** (1962), pp. 649–665.

[10] Deligne P., La conjecture de Weil I, *Publ. Math. IHES* **43** (1974), pp. 273–307.

[11] Heath-Brown D. R., Cubic forms in 10 variables, *Proc. London Math. Soc.* **47** (3) (1983), pp. 225–257.

[12] Katz, N., Sommes exponentielles, *Astérisque* **79**, Soc. Math. de France, 1980.

[13] Leep D., *Systems of quadratic forms* (to appear).

[14] Leep D. and Schmidt W. M., Systems of homogeneous equations, *Inventiones Math.* **71** (1983), pp. 539–549.

[15] Lewis D. J. and Montgomery H., On zeros of p-adic forms, *Michigan J. of Math.* **30** (1983), pp. 83–87.

[16] Myerson G., The distribution of rational points on varieties defined over a finite field, *Mathematika* **28** (1981), pp. 153–159.

[17] Schinzel A., Schlickewei H. P., and Schmidt W. M., Small solutions of quadratic congruences and small fractional parts of quadratic forms, *Acta Arith.* **37** (1980), pp. 241–248.
[18] Schmidt W. M., Diophantine inequalities for forms of odd degree, *Advances in Math.* **38** (1980), pp. 128–151.
[19] Schmidt W. M., Simultaneous p-adic zeros of quadratic forms, *Monatsh. Math.* **90** (1980), pp. 45–65.
[20] Schmidt W. M., On cubic polynomials, III. Systems of p-adic equations, *Monatsh. Math.* **93** (1982), pp. 211–223.
[21] Schmidt W. M., On cubic polynomials. IV. Systems of rational equations, *Monatsh. Math.* **93** (1982), pp. 329–348.
[22] Schmidt W. M., The solubility of certain p-adic equations, *J. of Number Theory* (to appear).
[23] Schmidt W. M., The density of integer points on homogeneous varieties, *Acta Math.* (to appear).
[24] Schmidt W. M., Some estimates of exponential sums, *Acta Arith.* (to appear).
[25] Serre J. P., Majorations de sommes exponentielles, *Astérisque* **41–44** (1977), pp. 111–126.
[26] Terjanian G., Un contre-exemple à une conjecture d'Artin, *C. R. Acad. Sci. Paris, Ser. A-B* **262** (1966), A 612.
[27] Terjanian G., Formes p-adiques anisotropes, *J. Reine Angew. Math.* **313** (1980), pp. 217–220.

UNIVERSITY OF COLORADO
BOULDER

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

J.-L. WALDSPURGER

Correspondances de Shimura

On donne ici un aperçu sur les articles [11] et [13], consacrés à l'étude des correspondances entre formes automorphes sur $\widetilde{\mathrm{SL}}_2$ et formes automorphes sur PGL_2 (ou une de ses formes tordues). On signale aussi quelques applications extraites de [10] et [12].

Le sujet remonte à Shimura ([8]). Les idées utilisées dans les travaux précités sont largement inspirées de l'article [9] de Shintani.

La démonstration des résultats ci-dessous est très sinueuse. Dans un souci de clarté, on les présente sans se soucier de l'ordre dans lequel ils sont démontrés.

1

Soient F un corps de nombres, $\boldsymbol{A}$ son anneau des adèles, $\boldsymbol{A}^\times$ son groupe des idèles. Si v est une place de F, on note F_v le complété de F en v.

1.1. Notons $\widetilde{\mathrm{SL}}_2(\boldsymbol{A})$ le groupe métaplectique, revêtement de degré 2 de $\mathrm{SL}_2(\boldsymbol{A})$ (cf. par exemple [2]), $\{1, \zeta\}$ le noyau de la projection $\widetilde{\mathrm{SL}}_2(\boldsymbol{A}) \to \mathrm{SL}_2(\boldsymbol{A})$. Il existe une section unique $\mathrm{SL}_2(F) \to \widetilde{\mathrm{SL}}_2(\boldsymbol{A})$. Identifions $\mathrm{SL}_2(F)$ et son image. Pour toute place v de F, on définit de même $\widetilde{\mathrm{SL}}_2(F_v)$.

Remarque. Le groupe métaplectique n'est pas un groupe algébrique.

Pour toute place v archimédienne de F, on choisit un sous-groupe compact maximal de $\widetilde{\mathrm{SL}}_2(F_v)$. On définit alors pour toute place v l'algèbre de Hecke $\tilde{\mathscr{H}}(F_v)$ de $\widetilde{\mathrm{SL}}_2(F_v)$, puis l'algèbre globale $\tilde{\mathscr{H}}(\boldsymbol{A})$. Soit $\tilde{\mathscr{A}}_0$ l'espace des formes automorphes paraboliques sur $\mathrm{SL}_2(F)\backslash\widetilde{\mathrm{SL}}_2(\boldsymbol{A})$, impaires (i.e. pour $f \in \tilde{\mathscr{A}}_0$ et $\sigma \in \widetilde{\mathrm{SL}}_2(\boldsymbol{A})$, on a $f(\zeta\sigma) = -f(\sigma)$). C'est un $\tilde{\mathscr{H}}(\boldsymbol{A})$-module, qui se décompose en somme directe de sous-modules irréductibles. On note un tel sous-module $(\tilde{\pi}, \tilde{E})$: $\tilde{E}$ est le sous-espace de $\tilde{\mathscr{A}}_0$ et $\tilde{\pi}$ la représentation de $\tilde{\mathscr{H}}(\boldsymbol{A})$ dans $\tilde{E}$. Si $(\tilde{\pi}, \tilde{E})$ est un tel sous-module, on peut définir pour toute place v de F une représentation admissible irréductible $\tilde{\pi}_v$ de $\tilde{\mathscr{H}}(F_v)$, de telle sorte que $\tilde{\pi} \simeq \bigotimes_v \tilde{\pi}_v$ (produit tensoriel restreint).

1.2. Soient M une algèbre de quaternions définie sur F, déployée ou non, $G = PM^\times$ (groupe adjoint du groupe des éléments inversibles de M). Pour toute place archimédienne v de F, on choisit un sous-groupe compact maximal de $G(F_v)$. On définit alors l'algèbre de Hecke $\mathscr{H}(G, \boldsymbol{A})$, et l'espace $\mathscr{A}_0(G)$ des formes automorphes paraboliques sur $G(F)\backslash G(\boldsymbol{A})$, orthogonales aux caractères. Ces objets vérifient des propriétés analogues à celles décrites ci-dessus (1.1). On adopte des notations analogues.

1.3. Soient $\psi\colon \boldsymbol{A}/F \to \boldsymbol{C}^\times$ un caractère continu, $\psi \neq 1$, H l'espace des éléments de M de trace nulle. Il est muni de la forme quadratique $q(x) = -N(x)$, où N est la norme (réduite) de M/F. Le groupe G agit sur H par conjugaison et s'identifie à $\mathrm{SO}(H, q)$. On définit alors une représentation de Weil ω_ψ de $\widetilde{\mathrm{SL}}_2(\boldsymbol{A}) \times G(\boldsymbol{A})$ dans l'espace de Schwartz $\mathscr{S}(H(\boldsymbol{A}))$ (cf. [14], ou [2]). On se restreint à un sous-espace $\mathscr{S}_\psi = \mathscr{S}_\psi(H(\boldsymbol{A}))$, en imposant des conditions aux places archimédiennes ([11], p. 4). Il n'est pas invariant par ω_ψ, mais l'est par la représentation de $\tilde{\mathscr{H}}(\boldsymbol{A}) \otimes \mathscr{H}(G, \boldsymbol{A})$ déduite de ω_ψ.

Si $f \in \mathscr{S}_\psi$, on définit une série thêta $\theta_{\psi,f}$ sur $\widetilde{\mathrm{SL}}_2(\boldsymbol{A}) \times G(\boldsymbol{A})$ par

$$\theta_{\psi,f}(\sigma, g) = \sum_{x \in H(F)} \omega_\psi(\sigma, g) f(x).$$

C'est une fonction invariante à gauche par $\mathrm{SL}_2(F) \times G(F)$.

1.4. De la même façon, on définit des séries thêta sur $\mathrm{SL}_2(F)\backslash \widetilde{\mathrm{SL}}_2(\boldsymbol{A})$ associées à l'espace F muni de la forme quadratique $q(x) = x^2$. Ce sont des formes automorphes. On note $\tilde{\mathscr{A}}_{00}$ l'espace des éléments de $\tilde{\mathscr{A}}_0$ orthogonaux (pour le produit de Petersson usuel) à toutes les séries thêta construites ainsi (en faisant varier ψ). C'est un $\tilde{\mathscr{H}}(\boldsymbol{A})$-module.

1.5. Notons $\tilde{A}_{00}$, resp. $A_0(G)$, l'ensemble des sous-modules irréductibles de $\tilde{\mathscr{A}}_{00}$, resp. $\mathscr{A}_0(G)$. On note $\mathrm{JL}\colon A_0(G) \to A_0(\mathrm{PGL}_2)$ la correspondance de Jacquet et Langlands ([4], § 15). On note de même ses analogues locales.

2

Revenons à la situation du 1.3.

2.1. Soit $(\pi, E) \in A_0(G)$. Pour $\varphi \in E$, $f \in \mathscr{S}_\psi$, $\sigma \in \widetilde{\mathrm{SL}}_2(A)$, posons

$$\theta_{\psi,f,\varphi}(\sigma) = \int_{G(F)\backslash G(A)} \theta_{\psi,f}(\sigma, g)\varphi(g)\,dg,$$

où dg est une mesure de Haar fixée. La fonction $\theta_{\psi,f,\varphi}$ appartient à $\tilde{\mathscr{A}}_0$. Notons $\Theta(E, \psi)$ l'espace des fonctions $\theta_{\psi,f,\varphi}$ quand f décrit $\mathscr{S}_\psi$, et φ décrit E.

PROPOSITION 1. *Si $\Theta(E,\psi)$ est non nul, c'est un sous-module irréductible de $\tilde{\mathscr{A}}_{00}$.* (Cf. [11], p. 80; [13], prop. 20).

2.2. Soit $(\tilde{\pi}, \tilde{E}) \in \tilde{A}_{00}$. Pour $\tilde{\varphi} \in \tilde{E}$, $f \in \mathscr{S}_\psi$, $g \in G(A)$, posons

$$\theta_{\psi,f,\tilde{\varphi}}(g) = \int_{\mathrm{SL}_2(F)\backslash \mathrm{SL}_2(A)} \overline{\theta_{\psi,f}(\sigma, g)}\tilde{\varphi}(\sigma)\,d\sigma.$$

La fonction $\theta_{\psi,f,\tilde{\varphi}}$ appartient à $\mathscr{A}_0(G)$. Notons $\Theta(\tilde{E}, \psi, G)$ l'espace des fonctions $\theta_{\psi,f,\tilde{\varphi}}$ quand f décrit $\mathscr{S}_\psi$ et $\tilde{\varphi}$ décrit $\tilde{E}$.

PROPOSITION 2. *Si $\Theta(\tilde{E},\psi,G)$ est non nul, c'est un sous-module irréductible de $\mathscr{A}_0(G)$.* (Cf. [11], p. 98; [13], prop. 20).

3

Les applications $E \mapsto \Theta(E, \psi)$ et $\tilde{E} \mapsto \Theta(\tilde{E}, \psi, G)$ sont de nature locale. Décrivons la situation locale.

3.1. Soit v une place finie de F. On note ici G' le groupe adjoint du groupe de quaternions non déployé sur F_v, $A_v(G')$ (resp. $A_v(\mathrm{PGL}_2)$) l'ensemble des classes d'isomorphie de représentations admissibles irréductibles de G' (resp. et de dimension infinie de $\mathrm{PGL}_2(F_v)$). Notons $\tilde{A}_v$ l'ensemble des classes d'isomorphie de repésentations $\tilde{\pi}_v$ admissibles irréductibles de $\widetilde{\mathrm{SL}}_2(F_v)$ telles que: (a) $\tilde{\pi}_v$ ne se factorise pas par $\mathrm{SL}_2(F_v)$, i.e. $\tilde{\pi}_v(\zeta) = -\mathrm{id}$, (b) $\tilde{\pi}_v$ n'est pas une représentation de Weil "élémentaire" (analogue p-adique des représentations de la série discrète holomorphe de poids $1/2$, ou antiholomorphe de poids $-1/2$, du cas réel). Soit ψ_v: $F_v \to C^\times$ un caractère continu, $\psi_v \neq 1$. On définit comme d'habitude la notion de modèle de Whittaker de $\tilde{\pi}_v \in \tilde{A}_v$, relativement à ψ_v. Un tel modèle est unique s'il existe, mais n'existe pas toujours. Notons $\tilde{A}_v(\psi_v)$, resp. $\tilde{A}'_v(\psi_v)$, l'ensemble des $\tilde{\pi}_v \in \tilde{A}_v$ pour lesquelles ce modèle existe, resp. n'existe pas.

3.2. Par une construction locale analogue à celle du n° 2, on définit des applications

$$A_v(\mathrm{PGL}_2) \xrightarrow{\theta(\cdot,\psi_v)} \tilde{A}_v(\psi_v), \quad \tilde{A}_v(\psi_v) \xrightarrow{\theta(\cdot,\psi_v,\mathrm{PGL}_2)} A_v(\mathrm{PGL}_2).$$

Ce sont des bijections réciproques. De même, on définit des applications

$$A_v(G') \xrightarrow{\theta(\cdot,\psi_v)} \tilde{A}'_v(\psi_v), \qquad \tilde{A}'_v(\psi_v) \xrightarrow{\theta(\cdot,\psi_v,G')} A_v(G').$$

Ce sont des bijections réciproques (cf. [13], th. 1 et prop. 14, et l'article fondamental [7]).

3.3. Via les bijections ci-dessus, la correspondance de Jacquet et Langlands définit une injection

$$R(\cdot, \psi_v): \tilde{A}'_v(\psi_v) \to \tilde{A}_v(\psi_v).$$

Notons $\mathscr{R}$ la relation d'équivalence dans $\tilde{A}_v$ qui rend équivalentes $\tilde{\pi}_v \in \tilde{A}'_v(\psi_v)$ et son image $R(\tilde{\pi}_v, \psi_v)$. Elle est indépendante de ψ_v. Les classes d'équivalence sont de la forme $\{\tilde{\pi}_v\}$ avec $\tilde{\pi}_v \in \tilde{A}_v(\psi_v) - \mathrm{Im}\, R(\cdot, \psi_v)$, ou $\{\tilde{\pi}_v^1, \tilde{\pi}_v^2\}$, avec $\tilde{\pi}_v^1 \in \tilde{A}'_v(\psi_v)$ et $\tilde{\pi}_v^2 = R(\tilde{\pi}_v^1, \psi_v)$. On note dans ce cas $\tilde{\pi}_v^1 = \mathscr{R}\tilde{\pi}_v^2$, $\tilde{\pi}_v^2 = \mathscr{R}\tilde{\pi}_v^1$. Ces deux représentations ont des caractères centraux différents (le centre de $\widetilde{\mathrm{SL}}_2(F_v)$ est l'image réciproque de celui de $\mathrm{SL}_2(F_v)$, il a 4 éléments).

3.4. On définit une application

$$\theta^0(\cdot, \psi_v): \tilde{A}_v \to A_v(\mathrm{PGL}_2),$$

égale à $\theta(\cdot, \psi_v, \mathrm{PGL}_2)$ sur $\tilde{A}_v(\psi_v)$ et à $\mathrm{JL} \circ \theta(\cdot, \psi_v, G')$ sur $\tilde{A}'_v(\psi_v)$.

3.5. Des résultats analogues sont vrais en une place v archimédienne, en considérant des $\mathfrak{g}$-K-modules, au lieu de représentations des groupes.

4

Revenons à la situation globale.

4.1. Soient $(\pi, E) \in A_0(G)$, $L(\pi, s)$ la fonction L habituelle de π, $\varepsilon(\pi, s)$ son facteur ε. Si $\Theta(E, \psi) \neq \{0\}$, notons $\theta(E, \psi)$ la représentation de $\tilde{\mathscr{H}}(A)$ dans cet espace. Notons $\theta(\pi, \psi)$ la représentation $\bigotimes_v \theta(\pi_v, \psi_v)$ de $\tilde{\mathscr{H}}(A)$ (cf. 3.2).

THÉORÈME 1. (1) $\Theta(E,\psi) \neq \{0\}$ *si et seulement si* $L(\pi,1/2) \neq 0$.

(2) *Si* $\Theta(E,\psi) \neq \{0\}$, $\theta(E,\psi)$ *et* $\theta(\pi,\psi)$ *sont isomorphes.*

(3) *Il existe* $(\tilde{\pi},\tilde{E}) \in \tilde{A}_{00}$ *tel que* $\tilde{\pi}$ *et* $\theta(\pi,\psi)$ *soient isomorphes, si et seulement si* $\varepsilon(\pi,1/2) = 1$. (Cf. [11], th. 1; [13], prop. 22; [6], th. A1).

4.2. Soit $(\tilde{\pi},\tilde{E}) \in \tilde{A}_{00}$. Si $\Theta(\tilde{E},\psi,G) \neq \{0\}$, notons $\theta(\tilde{E},\psi,G)$ la représentation de $\mathscr{H}(G,\boldsymbol{A})$ dans cet espace. Notons $\theta^0(\tilde{\pi},\psi)$ la représentation $\bigotimes_v \theta^0(\tilde{\pi}_v,\psi_v)$ de $\mathscr{H}(\mathrm{PGL}_2,\boldsymbol{A})$ (cf. 3.4), et $L(\tilde{\pi},\psi,s)$ sa fonction L (elle est "automorphe" en vertu du (3) ci-dessous).

THÉORÈME 2. (1) $\Theta(\tilde{E},\psi,G) \neq \{0\}$ *si et seulement si les trois conditions suivantes sont vérifiées:*

(i) $L(\tilde{\pi},\psi,1/2) \neq 0$,

(ii) *pour toute place* v *de* F *où* $M(F_v)$ *est déployée,* $\tilde{\pi}_v \in \tilde{A}_v(\psi_v)$,

(iii) *pour toute place* v *de* F *où* $M(F_v)$ *n'est pas déployée,* $\tilde{\pi}_v \in \tilde{A}'_v(\psi_v)$.

(2) *Si* $\Theta(\tilde{E},\psi,G) \neq \{0\}$, *les représentations* $\theta(\tilde{E},\psi,G)$ *et* $\bigotimes_v \theta\big(\tilde{\pi}_v,\psi_v,G(F_v)\big)$ *sont isomorphes, et* $\mathrm{JL}\big(\theta(\tilde{E},\psi,G)\big) = \theta^0(\tilde{\pi},\psi)$.

(3) *Il existe* $(\pi,E) \in A_0(\mathrm{PGL}_2)$ *tel que* π *et* $\theta^0(\tilde{\pi},\psi)$ *soient isomorphes.* (Cf. [13], prop. 21. La démonstration de ces théorèmes utilise les résultats de Flicker [1]).

4.3. Les applications $E \mapsto \Theta(E,\psi)$ et $\tilde{E} \mapsto \Theta(\tilde{E},\psi,G)$ sont des bijections réciproques entre les sous-ensembles de $A_0(G)$ et $\tilde{A}_{00}$ où elles ne s'annulent pas. On note que si $G \neq \mathrm{PGL}_2$, le diagramme suivant *n'est pas* commutatif

$$\begin{tikzcd} A_0(G) \arrow[d, "\mathrm{JL}"'] \arrow[dr, "\Theta(\cdot,\psi)"] & \\ A_0(\mathrm{PGL}_2) \arrow[r, "\Theta(\cdot,\psi)"'] & \tilde{A}_{00} \end{tikzcd}$$

(les applications Θ sont restreintes aux sous-ensembles ci-dessus...)

5

On donne ici des applications des résultats précédents à l'étude de $\tilde{A}_{00}$.

5.1. THÉORÈME 3 (*de multiplicité* 1). *Soient* $(\tilde{\pi}_i,\tilde{E}_i) \in \tilde{A}_{00}$, *pour* $i = 1, 2$. *Supposons* $\tilde{\pi}_1$ *et* $\tilde{\pi}_2$ *isomorphes. Alors* $\tilde{E}_1 = \tilde{E}_2$. ([11], p. 131).

5.2. Soient $(\tilde{\pi}_i, \tilde{E}_i) \in \tilde{A}_{00}$, pour $i = 1, 2$. On dit que $\tilde{\pi}_1$ et $\tilde{\pi}_2$ sont proches si $\tilde{\pi}_{1,v} \simeq \tilde{\pi}_{2,v}$ pour presque toute place v de F (cf. [3], § 6). Cela définit dans $\tilde{A}_{00}$ une relation d'équivalence, dont on peut décrire les classes. Pour $(\tilde{\pi}, \tilde{E}) \in \tilde{A}_{00}$, notons $\Sigma_{\tilde{\pi}}$ l'ensemble des places v de F telles que la classe d'équivalence de $\tilde{\pi}_v$ (pour la relation $\mathscr{R}$, cf. 3.3) a 2 éléments. C'est aussi l'ensemble des places v telles que $\theta^0(\tilde{\pi}_v, \psi_v)$ est dans l'image de l'application JL, où ψ_v est un caractère quelconque. L'ensemble $\Sigma_{\tilde{\pi}}$ est fini.

THÉORÈME 4. *Soit* $(\tilde{\pi}, \tilde{E}) \in \tilde{A}_{00}$.

(1) *Soit* $(\tilde{\pi}', \tilde{E}') \in \tilde{A}_{00}$. *Supposons* $\tilde{\pi}$ *et* $\tilde{\pi}'$ *proches. Alors il existe un sous-ensemble* $\Sigma \subset \Sigma_{\tilde{\pi}}$ *tel que*

(i) Σ *a un nombre pair d'éléments,*

(ii) *si* v *est une place de* F, *on a* $\tilde{\pi}'_v \simeq \tilde{\pi}_v$ *si* $v \notin \Sigma$, $\tilde{\pi}'_v \simeq \mathscr{R}\tilde{\pi}_v$ *si* $v \in \Sigma$.

(2) *Soit* $\Sigma \subset \Sigma_{\tilde{\pi}}$, *vérifiant* (i). *Il existe* $(\tilde{\pi}', \tilde{E}') \in \tilde{A}_{00}$ *vérifiant* (ii) ([13], cor. 2 à la prop. 19).

6

On donne ici des applications arithmétiques des résultats précédents. Le point central est le (1) du théorème 1, plus exactement sa démonstration qui utilise le calcul des coefficients de Fourier des fonctions $\theta_{\psi,f,\varphi}$.

6.1. THÉORÈME 5. *Soient* $(\pi, E) \in A_0(\mathrm{PGL}_2)$, Σ *un ensemble fini de places de* F. *Supposons* $\varepsilon(\pi, 1/2) = 1$. *Alors il existe un caractère quadratique* χ *de* $F^\times \setminus A^\times$ *tel que*

(i) $L(\pi \otimes \chi, 1/2) \neq 0$,

(ii) *pour toute place* $v \in \Sigma$, $\chi_v = 1$ ([13], th. 4).

6.2. Supposons désormais $F = Q$. On peut calculer les coefficients de Fourier des formes modulaires holomorphes de poids demi-entier. Des valeurs de fonctions L interviennent dans les formules obtenues (cf. [12] et aussi [5] qui utilise d'autres méthodes).

6.3. Vignéras en a déduit le théorème de rationalité suivant. Notons ∞ la place réelle de Q. Soit $(\pi, E) \in A_0(\mathrm{PGL}_2)$, tel que π_∞ soit de la série discrète de poids k (k pair $\geqslant 2$). Notons $Q(\pi)$ le sous-corps de C engendré par les valeurs propres des opérateurs de Hecke T_p agissant dans E, pour presque tout p. C'est une extension finie de Q totalement réelle. Soit S_π la réunion de $\{\infty\}$ et de l'ensemble des places finies p de Q telles que

π_p soit ramifiée (i.e. n'admet pas de vecteur invariant par le sous-groupe compact maximal standard de $\mathrm{PGL}_2(\boldsymbol{Q}_p)$). Si χ est un caractère quadratique de $\boldsymbol{Q}^\times\backslash\boldsymbol{A}^\times$, on note D_χ le discriminant de l'extension quadratique de $\boldsymbol{Q}$ associée à χ.

THÉORÈME 6. *Soient χ_1, χ_2 deux caractères quadratiques de $\boldsymbol{Q}^\times\backslash\boldsymbol{A}^\times$. Supposons*

(i) *$\chi_{1,v} = \chi_{2,v}$ pour toute place $v \in S_\pi$,*

(ii) *$L(\pi\otimes\chi_2, 1/2) \neq 0$.*

Alors il existe $q \in \boldsymbol{Q}(\pi)$ tel que

$$L(\pi\otimes\chi_1, 1/2)D_{\chi_1}^{(k-1)/2} = q^2 L(\pi\otimes\chi_2, 1/2)D_{\chi_2}^{(k-1)/2}$$

([10], p. 333).

Bibliographie

[1] Flicker Y., Automorphic Forms on Covering Groups of GL(2), *Invent. Math.* **57** (1980), pp. 119–182.

[2] Gelbart S., *Weil's Representation and the Spectrum of the Metaplectic Group*, Springer Lecture Notes **530**, 1976.

[3] Gelbart S. and Piatetski-Shapiro I. I., *Some Remarks on Metaplectic Cusp Forms and the Correspondences of Shimura and Waldspurger*, preprint.

[4] Jacquet H. and Langlands R. P., *Automorphic Forms on* GL(2), Springer Lecture Notes **114**, 1970.

[5] Kohnen W. and Zagier D., Values of L-series of Modular Forms at the Center of the Critical Strip, *Invent. Math.* **64** (1981), pp. 175–198.

[6] Piatetski-Shapiro I. I., Notes d'un cours à Maryland, à paraitre au *Israel Jour. of Math.*

[7] Rallis S. and Schiffmann G., Représentations supercuspidales du groupe métaplectique, *J. Math. Kyoto Univ.* **17** (1977), pp. 567–603.

[8] Shimura G., Modular Forms of Half Integral Weight, in: *Modular Functions of One Variable I*, Springer Lecture Notes **320**, 1973.

[9] Shintani T., On Construction of Holomorphic Cusp Forms of Half Integral Weight, *Nagoya Math. J.* **58** (1975), pp. 83–126.

[10] Vignéras M. F., Valeur au centre de symétrie des fonctions L associées aux formes modulaires, in: *Séminaire de Théorie des Nombres, Paris 1979–80*, Progress in Math. **12**, Birkhäuser, 1981.

[11] Waldspurger J.-L., Correspondance de Shimura, *J. Math. Pures Appl.* **59** (1980), pp. 1–133.

[12] Waldspurger J.-L., Sur les coefficients de Fourier des formes modulaires de poids demi-entier, *J. Math. Pures Appl.* **60** (1981), pp. 375-484.

[13] Waldspurger J.-L., *Correspondances de Shimura et quaternions*, preprint.

[14] Weil A., Sur certains groupes d'operateurs unitaires, *Acta Math.* **111** (1964), pp. 143–211.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

S. Y. CHENG*

On the Real and Complex Monge–Ampère Equation and Its Geometric Applications

§ 0. Introduction

Let u be a real-valued C^2 function on an open set in $\boldsymbol{R}^n$ or $\boldsymbol{C}^n$. Then we can form its Hessian matrix, $\left[\frac{\partial^2 u}{\partial x^i \partial x^j}\right]$ in the real case and $\left[\frac{\partial^2 u}{\partial z^i \partial \bar{z}^j}\right]$ in the complex case. $\left[\frac{\partial^2 u}{\partial x^i \partial x^j}\right]$ is a symmetric matrix and $\left[\frac{\partial^2 u}{\partial z^i \partial \bar{z}^j}\right]$ is Hermitian symmetric. The Monge–Ampère operator, denoted by $M(u)$, is the determinant of the Hessian matrix. An equation of the form $M(u) = F(x, u, \nabla u)$ is a nonlinear second order partial differential equation. It is called *elliptic* if the Hessian matrix of u is positive definite. This means that u is strictly convex in the real case and strictly plurisubharmonic in the complex case. Monge–Ampère equations of elliptic type arise naturally from geometric problems. In what follows we describe some of these problems.

§ 1. Real Monge–Ampère equation

One of the classical geometric problems which give rise to a real Monge–Ampère equation is the Minkowski problem about recovering an ovaloid from its Gauss–Kronecker curvature. This problem was studied and solved by H. Lewy, L. Nirenberg and A. Pogorelov. The reader may consult [10] for an exposition of this problem.

Affine differential geometry is another source from which Monge–Ampère equations arise naturally. Roughly speaking, affine differential

* Research partially supported by an NSF grant.

geometry consists in studying properties of convex hypersurfaces invariant under unimodular transformations and translations. In affine differential geometry there is no distinction between an ellipsoid and a sphere. Let M be a piece of a strictly convex hypersurface in $\boldsymbol{R}^{n+1}$. Then one can define a Riemannian metric, called the affine metric on M, invariantly. To express it explicitly, suppose that $M = \{(x, u(x))\}$ is the graph of a function u. Then the *affine metric on* M is given by $ds^2 = \sum \big(\det(u_{ij})\big)^{-1/(n+2)} u_{ij} dx^i dx^j$. One can also define a normal, called the *affine normal*, invariantly. Let M be the boundary of a convex body $\tilde{M}$ and let $x_0 \in M$. The affine normal of M at x_0 is the tangent vector to the curve described by the center of gravity of the intersections of $\tilde{M}$ with parallel translates of the tangent plane of M at x_0. If all the affine normals of M meet at one point or at infinity then M is called an *affine sphere*. There are three possible cases. The first case in the elliptic case and the model is the ellipsoid. In this case all the affine normals meet at the convex side of the hypersurface. The second case is the parabolic case and the model is the paraboloid. In this case all the affine normals meet at infinity, i.e. are parallel to each other. The third case is the hyperbolic case. In this case all the affine normals meet at the concave side. The hyperboloid is an example. Naturally one would like to conjecture that a strictly convex hypersurface which is closed w.r.t. the Euclidean topology and an affine sphere must be a hyperquadric. In order to show this, a cubic tensor called the Fubini–Pick form is very useful. If the hypersurface is the graph of a function f then its Fubini–Pick form is essentially composed of the 3rd derivatives of f. The Fubini–Pick form vanishes identically iff the hypersurface is a hyperquadric. The length of the Fubini–Pick form satisfies a very useful differential inequality. Let M be an affine sphere such that S is the length of the Fubini–Pick form and Δ is the Laplacian w.r.t. the affine metric. E. Calabi [1] computed ΔS and obtained that $\Delta S \geqslant (n+1)\cdot(S^3+HS)$, where H is positive, zero or negative according as M is an elliptic, parabolic or hyperbolic affine sphere. A local computation also shows that the Ricci curvature of affine spheres are bounded from below. Thus if M is an elliptic or parabolic affine sphere which is complete w.r.t. its affine metric then $S \equiv 0$, i.e. M is a hyperquadric. Now it remains to show that if M is an elliptic or a parabolic affine sphere which is closed w.r.t. the Euclidean topology then M is complete w.r.t. the affine metric. The elliptic case is solved by E. Calabi. In the parabolic case Calabi proved this for $n \leqslant 5$ and A. V. Pogorelov proved this for general n. In [3] a different proof of completeness is given by a gradient estimate. The case of hyperbolic affine sphere is different. E. Calabi [1] constructed examples of hyperbolic affine spheres which are not hyperquadrics. For example, any component of $x^1 \ldots x^{n+1} = 1$ in $\boldsymbol{R}^{n+1}$ is a hyperbolic affine sphere. E. Calabi [1] then proposed to prove that

(i) Every complete hyperbolic affine sphere is asymptotic to a convex cone.

(ii) Given any convex cone there is a hyperbolic affine sphere asymptotic to it.

THEOREM 1.1 [3]. *Every complete hyperbolic affine sphere is asymptotic to a convex cone.*

(ii) is equivalent to the following boundary value problem of a real Monge–Ampère equation on a bounded convex domain Ω

$$\begin{gathered} \det(u_{ij}) = \left(-\frac{1}{u}\right)^{n+2} \quad \text{on } \Omega, \\ u = 0 \quad \text{on } \partial\Omega, \\ (u_{ij}) > 0. \end{gathered} \tag{1.1}$$

Solution of (1.1) follows from the paper [4] concerning the solution of general boundary value problems for equations of the form $\det(u_{ij}) = F(x, u)$.

THEOREM 1.2 [4]. *There is a unique solution of* (1.1) *in* $C^\infty(\Omega) \cap C^0(\bar{\Omega})$.

Equation (1.1) also arises from another geometric problem. This is the problem of construction of a Riemannian metric on Ω invariant under projective transformations. Loewner and Nirenberg [9] showed that, if u is a solution of equation (1.1) then $\sum\left(\frac{u_{ij}}{-u}\right)dx^i dx^j$ is such a metric. To study the properties of u near the boundary we can use the following

THEOREM 1.3. *Let Ω be a bounded strictly convex domain with smooth boundary. Then there exists $u \in C^\infty(\Omega) \cap C^0(\bar{\Omega})$ satisfying* (1.1) *and such that $\sum\left(\frac{u_{ij}}{-u}\right)dx^i dx^j$ defines a complete metric on Ω which is invariant under projective transformations and its sectional curvatures are asymptotically equal to -1 near $\partial\Omega$.*

So far, the real Monge–Ampère equation arises from extrinsic geometry. It also arises naturally if M has some special geometric structure. One example is that M has an affine flat structure, i.e. M is a differentiable manifold which admits an atlas of coordinate charts such that the change of coordinates are given by affine transformations. In [6], a metric ds^2 on an affine flat manifold M is called a *Kähler affine metric* if on each affine flat coordinate charts $(x^1, \ldots, x^n)$ there exists a convex function f such that $ds^2 = \sum \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i dx^j$. In this case M is called a *Kähler affine flat manifold.* Then the techniques of parabolic affine spheres can be generalized, leading to the following

THEOREM 1.4 [6]. *Suppose that M is compact affine flat manifold which carries a parallel volume. Then M admits a Riemannian flat structure.*

§ 2. Complex Monge–Ampère equation

The complex Monge–Ampère operator $M(u)$ is defined by $M(u) = \det\left[\frac{\partial^2 u}{\partial z^i \partial \bar{z}^j}\right]$. In the complex case $M(u)$ is a very natural operator, in the sense that if $(z^{*1}, \ldots, z^{*n})$ is another set of analytic coordinates then

$$M(u) = \det\left[\frac{\partial^2 u}{\partial z^i \partial \bar{z}^j}\right] = \left|\frac{\partial z^{*i}}{\partial z^j}\right|^2 \det\left[\frac{\partial^2 u}{\partial z^{*i} \partial \bar{z}^{*j}}\right].$$

Therefore, a deep understanding of $M(u)$ will be very useful in complex analysis. In the case of compact Kähler manifolds, S. T. Yau [11] showed that solving complex Monge–Ampère equations on compact Kähler manifolds has many important consequences in differential geometry and algebraic geometry. Let us only quote one of the results in [11] which is relevant to our discussions here.

THEOREM 2.1 [11] (S. T. Yau). *Suppose that M is a compact Kähler manifold with ample cannonical class. Then,*

$$(-1)^n c_1^{n-2}(M) \cdot c_2(M) \geqslant (-1)^n \frac{n}{2(n+1)} c_1^n(M), \tag{2.1}$$

where $n = \dim_C M$. Equality in the above inequality holds iff M is covered by the ball biholomorphically.

The method of the above theorem is to show that the assumption on the canonical class gives rise to a Kähler metric which can be deformed to an Einstein–Kähler metric by solving a complex Monge–Ampère equation. Inequality (2.1) then follows from representing Chern classes by the curvatures of the Einstein–Kähler metric.

In the rest of this section we will discuss the problem of constructing complete Einstein–Kähler metrics on noncompact Kähler manifolds.

Let Ω be a bounded smooth strictly pseudoconvex domain. Let φ be a defining function of Ω, i.e. such that $\Omega = \{\varphi < 0\}$, $d\varphi \neq 0$ on $\partial\Omega$ and $(\varphi_{i\bar{j}}) > 0$ on $\bar{\Omega}$. Then $\sum g_{i\bar{j}}\, dz^i d\bar{z}^j = \sum \frac{\partial^2(-\log(-\varphi))}{\partial z^i \partial \bar{z}^j} dz^i d\bar{z}^j$ defines a complete Kähler metric on Ω. A computation shows that the holomorphic sectional curvatures of this metric is asymptotically equal to -1 near $\partial\Omega$. Also $\det[g_{i\bar{j}}] = e^{(n+1)g} e^F$ for some $F \in C^\infty(\bar{\Omega})$. Clearly, this metric is close to Kähler–Einstein near $\partial\Omega$. All we have to do is to perturb this metric to an Einstein–Kähler one. This follows from solving a Monge–Ampère equation on Ω of the form

$$\det[g_{i\bar{j}} + u_{i\bar{j}}] = e^{(n+1)u} e^{-F} \det[g_{i\bar{j}}], \tag{2.2}$$

such that $A^{-1}(g_{i\bar{j}}) < (g_{i\bar{j}} + u_{i\bar{j}}) < A(g_{i\bar{j}})$ for some positive constant A. Also notice that $v = -\log(-\varphi) + u$ satisfies

$$\begin{gathered} \det[v_{i\bar{j}}] = e^{(n+1)v}, \\ v = \infty \quad \text{on } \partial\Omega, \\ (v_{i\bar{j}}) > 0. \end{gathered} \tag{2.3}$$

C. Fefferman [7] showed that (2.3) is equivalent to

$$\begin{gathered} J(\psi) = (-1)^n \det \begin{bmatrix} \psi & \psi_{\bar{j}} \\ \psi_i & \psi_{i\bar{j}} \end{bmatrix} = 1, \\ \psi = 0 \quad \text{on } \partial\Omega, \end{gathered} \tag{2.4}$$

if $\psi = e^{-u}$. He constructed an approximate solution ψ to (2.3) in the sense that $J(\psi) = 1 + O(\psi^s)$, $s \leqslant n+2$. His work on the boundary behaviour of Bergman kernel shows that, in general, ψ cannot be more than $n+2$ times differentiable on $\bar{\Omega}$. In [5] we proved that there exists $\psi \in C^\infty(\Omega) \cap C^{n+3/2-\delta}(\bar{\Omega})$ satisfying (2.4) for any $\delta \in (0,1)$. Lee and Melrose [8] showed by using different methods that $\psi \in C^\infty(\Omega) \cap C^{n+2-\delta}(\bar{\Omega})$, $\delta \in (0,1)$. Summarizing, we have

THEOREM 2.2. *Let Ω be a bounded pseudoconvex domain in C^n. Then there exists a unique solution v satisfying equation* (2.3) *such that $\sum v_{i\bar{j}}\, dz^i d\bar{z}^j$ defines a complete Einstein–Kähler metric on Ω which is invariant under biholomorphic transformations. When Ω is assumed to be smooth and strictly pseudoconvex, then $e^{-v} \in C^\infty(\Omega) \cap C^{n+2-\delta}(\bar{\Omega})$ for $\delta \in (0,1)$ and the holo-*

morphic sectional curvature of $\sum v_{i\bar{j}}dz^i d\bar{z}^j$ *is asymptotically equal to* -1 *near* $\partial\Omega$.

In the above case the Chern forms of the Einstein–Kähler metric are not integrable and so we cannot expect to obtain results similar to Theorem 2.1. We have to restrict ourselves to the case when the infinity is rather small. The model for this case is the punctured disc and the Poincaré metric $|dz|^2/|z|^2(\log|z|^2)^2$. This direction is motivated by the work of P. Griffiths and J. Carson [2]. Let V be a compact Kähler manifold and D be a divisor with normal crossing such that $c_1(K_V)+c_1(L)>0$, where L is the line bundle corresponding to D. Griffiths and Carson then constructed a volume form on V to study holomorphic maps into V. We can use similar methods as before to construct a complete Einstein–Kähler metric on $V\setminus D$. The Chern forms of this Einstein–Kähler metric are integrable. Its integral represents logarithmic Chern classes $c_i(\log D)$ of D, i.e. Chern classes of the cotangent bundle over V whose sections are given by

$$\sum_{i=1}^{k} a_i(z)\frac{dz^i}{z^i}+\sum_{i=k+1}^{n} b_i(z)\,dz^i \quad \text{when } D \text{ is given by } \prod_{i=1}^{k} z^i = 0.$$

So we get an inequality similar to (2.1). In this case, if an equality holds, it still does imply constant curvature. However, we know explicitly that this is impossible if $c_1(K_V)+c_1(L)>0$ on V. We know that actually $c_1(K_V)+c_1(L)$ vanishes on D. However if we assume $c_1(L)<0$ on a neighborhood of D then we can still have a complete metric which we can perturb to an Einstein–Kähler metric.

Summing up we have

THEOREM 2.3. *Let V be a compact Kähler manifold with complex dimension n and let D be a divisor with normal crossing. Let L be the line bundle corresponding to D. We assume that either*

(i) $c_1(K_V)+c_1(L)>0$ *on* V

or

(ii) $c_1(K_V)+c_1(L)>0$ *on* $V\setminus D$, $c_1(K_V)+c_1(L)=0$ *on* D,

and that $c_1(L)<0$ *in a neighborhood of* D.

Then there exists a complete Einstein–Kähler metric on $V\setminus D$ which is invariant under biholomorphic maps. Moreover, for the logarithmic Chern classes of D we have

$$c_1^{n-2}(\log D)\cdot c_2(\log D) \geqslant \frac{n}{2(n+1)}\, c_1^n(\log D),$$

equality holding iff $V\setminus D$ is a quotient of the ball.

§ 3. Boundary value problems

In our approach to the solution of Monge–Ampère equations in the previous sections we use interior estimates. Recently, L. Caffarelli, J. J. Kohn, L. Nirenberg and J. Spruck obtained the boundary estimates and hence solved the boundary value problem. This will have important applications to nonlinear partial differential equations and geometric problems related to Monge–Ampère equations.

References

[1] Calabi E., *Complete Affine Hyperspheres I*, Istituto Nazionale di Alta Matematica, Symposia Mathematica, Volume X (1972).

[2] Carson J. and Griffiths P., A Defect Relation for Equidimensional Holomorphic Mappings between Algebraic Varieties, *Annals of Math.* **95** (1972).

[3] Cheng S. Y. and Yau S. T., *Complete Affine Hypersurfaces*, Part I: *The Completeness of Affine Metrics.*

[4] Cheng S. Y. and Yau S. T., On the Regularity of the Monge–Ampère Equation $\det\left(\frac{\partial^2 u}{\partial x^i \partial x^j}\right) = F(x, u)$, *Comm. on Pure and Appl. Math.* **XXX** (1977).

[5] Cheng S. Y. and Yau S. T., On the Existence of a Complete Kähler Metric on Non-Compact Complex Manifolds and the Regularity of Fefferman's Equation, *Comm. on Pure and Appl. Math.* **XXXIII** (1980).

[6] Cheng S. Y. and Yau S. T., The Real Monge–Ampère Equation and Affine Flat Structures. In: *Proceedings of the 1980 Beijing Symposium on Differential Geometry and Differential Equations*, Gordon and Breach, Science Publishers, Inc., New York, 1982.

[7] Fefferman C., Monge–Ampère Equations, the Bergman Kernel, and Geometry of Pseudoconvex Domains, *Annals of Math.* **103** (1976).

[8] Lee J. and Melrose R., *Boundary Behavior of the Complex Monge–Ampère Equation* (preprint).

[9] Loewner C. and Nirenberg L., Partial Differential Equations Invariant under Conformal or Projective Transformations. In: *Contribution to Analysis*, Academic Press, New York, 1974.

[10] Pogorelov A. V., *The Minkowski Multidimensional Problem*, translated by V. Oliker, *Scripta Series in Mathematics*, V. H. Winston and Sons, Washington, D. C., 1978.

[11] Yau S. T., Calabi's Conjecture and Some New Results in Algebraic Geometry, *Proc. Nat. Acad. Sci. USA* **74**, No. 5, May 1977.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

N. J. HITCHIN

The Geometry of Monopoles

§ 1. Introduction

When, five or six years ago, pure mathematicians found themselves using geometrical techniques to construct solutions to the Yang–Mills equations, the emphasis was on "instantons". These were solutions in $\boldsymbol{R}^4$ with finite action, and they could be tackled by means of a conformal compactification of $\boldsymbol{R}^4$ to S^4. There is another class of solutions, however, the "magnetic monopoles", which resisted attack until more recently. In the past three years new techniques have been applied to construct solutions. These reveal basic geometric structures, linking up with nonlinear equations like the KdV equation which themselves have received a great deal of attention in recent years. The "magnetic monopoles" are solutions to the *Bogomolny equations*. The basic data consist of a principal G-bundle with connection on $\boldsymbol{R}^3$ and a section Φ of the associated bundle of Lie algebras: Φ is known as the *Higgs field*. The equations are

$$D\Phi = *F, \tag{1}$$

where D denotes the covariant derivative and F the curvature of the connection. These are particular solutions of the second order *Yang–Mills–Higgs* equations with zero potential term:

$$\begin{aligned} D^* D\Phi &= 0, \\ D^* F &= [D\Phi, \Phi]. \end{aligned} \tag{2}$$

The appropriate boundary conditions [8] imply that Φ should approach (in a suitable gauge) a constant value in radial directions as $r \to \infty$. This provides a reduction of the bundle at infinity to the stabilizer K of an element of the Lie algebra, and hence a map from the sphere at infinity to G/K. There is thus a topological invariant in $\pi_2(G/K)$ which is generally referred to as a "magnetic charge" [5].

If $G = \mathrm{U}(1)$, then the equation (1) becomes simply

$$\operatorname{grad}\phi = \operatorname{curl}\mathbf{A}.$$

The fact that $\mathbf{A}$ should be regarded as a connection form, and not as a vector field, is already latently evident in Dirac's paper of 1931 (see [15]), so the differential-geometric approach is well established.

The first nonabelian example, for $G = \mathrm{SU}(2)$, was found in 1975 ([2], [11]) and has spherical symmetry. The magnetic charge (an element of $\pi_2(\mathrm{SU}(2)/\mathrm{U}(1)) \cong \mathbf{Z}$) for this solution is $k = 1$. Solutions for higher charge were unknown until Jaffe and Taubes' existence theorem in 1980 [8]. Then quite rapidly Ward [14] and Forgács, Horváth and Palla [4] discovered explicitly a charge 2 solution, Prasad and Rossi [12] extended the method to axially symmetric higher charge solutions and Corrigan and Goddard [3] realized a $(4k-1)$-parameter family of solutions of charge k. The author [6] gave a geometrical treatment of Ward's twistor approach, but throughout problems of nonsingularity remained, even forcing recourse to computers.

A new viewpoint, due to Nahm [10], but based on the ADHM construction of instantons dealt well with questions of nonsingularity and reduced the equations to a system of nonlinear ordinary differential equations.

At the moment these two approaches play complementary roles, each in its own way adequate to describe the solution [7], but each emphasizing a different aspect. I shall describe them in this lecture. The methods of Forgács et al., relying on Bäcklund transformations, constitute a third and not unrelated approach.

§ 2. The twistor method

Ward's approach was based on the complex twistor geometry of Penrose. The Bogomolny equations may be considered as the self-duality equations in $\mathbf{R}^4$, which are in addition invariant under translation in the x_0-direction. It follows that a solution is determined by a holomorphic vector bundle $\tilde{E}$ on the quotient space of $\mathbf{CP}^3/\mathbf{CP}^1$ by an action of $\mathbf{C}$ induced from the translation.

The quotient space is naturally the tangent bundle $\mathbf{TP}^1$ to $\mathbf{CP}^1$ and geometrically may be thought of as the space of oriented straight lines in $\mathbf{R}^3$: to each straight line we associate the (unit) direction $\mathbf{u}$, and the vector $\mathbf{x}$ which realizes the shortest distance to the origin. Since $\mathbf{x}\cdot\mathbf{u} = 0$, $\mathbf{x}$ is a tangent vector to S^2 at $\mathbf{u}$ and we thus obtain $\mathbf{TP}^1$.

The direct method to define the vector bundle $\tilde{E}$ is to set

$$\tilde{E}_z = \{s \in \Gamma(\gamma_z, E) |\ (\nabla_u - i\Phi)s = 0\},$$

where γ_z is the straight line corresponding to $z \in TP^1$ and E is the rank 2 vector bundle on R^3 associated with the given SU(2) principal bundle. Thus the fibre at z is the space of solutions of an ordinary differential equation along the line.

This much requires no boundary conditions, but considering the asymptotic condition

$$\Phi = i\begin{bmatrix} 1-\dfrac{k}{2r} & 0 \\ 0 & -1+\dfrac{k}{2r} \end{bmatrix} + O(r^{-2})$$

one may find a distinguished 1-dimensional subspace $L_z^+ \subset \tilde{E}_z$ of solutions to $(\nabla_u - i\Phi)s = 0$ which asymptotically decay as $t \to +\infty$ like $e^{-t}t^{k/2}$. As z varies, these subspaces describe a holomorphic line bundle over TP^1 which is moreover isomorphic to $L(-k) = L \otimes O(-k)$, where L is a natural $SL(2, C)$-invariant line bundle on TP^1, and $O(k)$ the pull-back of the standard bundle of degree k on CP^1.

It follows that $\tilde{E}$ is an extension of line bundles and is thus given by an extension class in the sheaf cohomology group $H^1(TP^1, L^2(-2k))$.

There is also another line bundle $L^- \cong L^*(-k)$ consisting of solutions which decay as $t \to -\infty$, which means that $\tilde{E}$ is likewise expressible by an extension class in $H^1(TP^1, L^{-2}(-2k))$.

The two subbundles coincide over an algebraic curve $S \subset TP^1$, in the linear system $|O(2k)|$. Since $L^+ = L^-$ on S, there is the fundamental constraint $L_S^2 \cong O_S$ on S. The curve S is called the *spectral curve* of the monopole. In fact, it determines completely the holomorphic bundle $\tilde{E}$, and consequently the solution to the Bogomolny equations, via the coboundary map

$$C \cong H^0(S, L^2) \xrightarrow{\delta} H^1(TP^1, L^2(-2k)).$$

Thus every solution of charge k is determined by a curve in the linear system $|O(2k)|$. If we take coordinates η, ζ on TP^1 by

$$(\eta, \zeta) \mapsto \eta \frac{d}{d\zeta}$$

then S is a curve of the form

$$\eta^k + a_1(\zeta)\eta^{k-1} + \ldots + a_k(\zeta) = 0,$$

where $a_l(\zeta)$ is a polynomial of degree $2l$.

§ 3. Nahm's construction

To understand Nahm's construction, recall the ADHM construction of finite action self-dual SU(2)-connections on $\boldsymbol{R}^4$ as given, for example, in [1]. To construct a connection on S^4 with Chern class $-k$, one takes a $(k+1)\times k$ quaternionic matrix

$$\Delta(x) = Ax + B$$

depending linearly on the quaternion $x \in \boldsymbol{H} \cong \boldsymbol{R}^4$, and of maximal rank. Then Δ defines a map

$$f\colon \boldsymbol{H} \to \boldsymbol{HP}^k, \quad f(x) = \ker \Delta^*(x)$$

and the pull-back of the quaternionic Hopf bundle is self-dual if $\Delta^*\Delta$ is *real* for all $x \in \boldsymbol{H}$.

Nahm replaces this finite-dimensional matrix with an infinite-dimensional one — a differential operator

$$\Delta(x)f = xf + i\frac{df}{dz} + i\sum_{j=1}^{3} T_j(z)e_j f$$

acting on functions $f\colon [0, 2] \to \boldsymbol{C}^k \otimes \boldsymbol{C}^2$. The operators e_1, e_2, e_3 consist of left multiplication on the quaternions $\boldsymbol{C}^2$ by i, j, k and the T_j's are $k\times k$ matrices. If the T_j's satisfy suitable reality conditions, $\Delta(x)$ may be thought of as a quaternionic operator, and the condition that $\Delta^*\Delta$ is real gives:

$$\begin{aligned} T_j &= -T_j^*, \\ \frac{dT_1}{dz} &= [T_2, T_3], \\ \frac{dT_2}{dz} &= [T_3, T_1], \\ \frac{dT_3}{dz} &= [T_1, T_2]. \end{aligned} \tag{3}$$

These equations have themselves a considerable geometrical content. They occur, moreover, spontaneously in a number of areas of mathematics, for example in Euler's equations for a spinning top, or in Schmid's work on the variation in the period matrix of a degenerating family of algebraic varieties.

If one writes

$$\begin{aligned} A(\zeta) &= (T_1+iT_2)-2iT_3\zeta+(T_1-iT_2)\zeta^2, \\ A_+(\zeta) &= \qquad\qquad\quad -iT_3+(T_1-iT_2)\zeta \end{aligned}$$

for an indeterminate ζ, then Nahm's equations (3) become

$$\frac{dA}{dz} = [A, A_+].$$

It follows from this Lax form that

$$\det(\eta+A) = \eta^k+a_1(\zeta)\eta^{k-1}+\ldots+a_k(\zeta)$$

is independent of z, and $\det\big(\eta+A(\zeta)\big) = 0$ describes an algebraic curve S in TP^1. Not surprisingly, this is the spectral curve of the twistor construction [7].

As (η, ζ) vary on S, $\ker\big(\eta+A(\zeta)\big)$ describes a holomorphic line bundle on S. As z varies, the line bundle describes a curve in the Jacobian of S. The basic reinterpretation of Nahm's equation, in a context which is used to generate solutions to the KdV equations and many others, is that this curve is a straight line in the direction determined by the bundle L on TP^1. In order to obtain the correct boundary conditions for Nahm's construction, the curve must be periodic and pass through the origin. This is the condition $L_S^2 = O_S$.

Thus a single curve S determines in radically different ways a solution of the Bogomolny equations.

§ 4. Further developments

The twistor approach has been used by Murray [9] to analyse monopoles for arbitrary simple Lie groups G, where the reduction at infinity is to a maximal torus T. Since

$$\pi_2(G/T) \cong Z^l \quad (l = \operatorname{rank} G),$$

there are l magnetic charges $(k_1, \ldots, k_l)$. One obtains then l spectral curves defined by polynomials in η of degree k_i. There is one for each vertex of the Dynkin diagram for G, with compatibility conditions on their points of mutual intersection whenever the vertices are adjacent. The monopole is generically determined by the spectral curves by repeated extensions.

The spectral curve approach is also useful in analyzing instantons on $\boldsymbol{R}^3 \times S^1$, i.e., *periodic* in the x_0-variable. With appropriate boundary conditions these can be thought of as monopoles for the loop group of G, and the formalism of the general simple Lie group passes over, so that one finds an extra spectral curve corresponding to the highest root. In the case of SU(2), this now parametrizes the "jumping lines" for a bundle on a 3-dimensional twistor space. To recover the bundle from the curves requires Ext groups rather than simply cohomology groups, as in the monopole case.

Perhaps the most tantalizing question to be dealt with in the future is the construction of Taubes' *non-self-dual* SU(2) monopoles [13]. These satisfy the second order equations (2). Of all the analogous equations (Einstein, Yang–Mills, Yang–Mills–Higgs) these provide perhaps the greatest chance of finding geometrical constructions which are not restricted to self-duality.

References

[1] Atiyah M. F., *Geometry of Yang–Mills Fields* (Fermi lectures), Scuola Normale Superiore, Pisa, 1979.

[2] Bogomolny E. B., The Stability of Classical Solutions, *Soviet J. Nuclear Phys.* **24** (1976), pp. 449–454.

[3] Corrigan E. and Goddard P., An n Monopole Solution with $4n-1$ Degrees of Freedom, *Comm. Math. Phys.* **80** (1981), pp. 575–587.

[4] Forgács P., Horváth Z., and Palla L., Exact Multimonopole Solutions in the Bogomolny–Prasad–Sommerfield Limit, *Phys. Lett.* **102** B (1981), pp. 131–135.

[5] Goddard P. and Olive D. I., Magnetic Monopoles in Gauge Field Theories, *Rep. Progr. Phys.* **51** (1978), pp. 1361–1435.

[6] Hitchin N. J., Monopoles and Geodesics, *Comm. Math. Phys.* **83** (1982), pp. 579–602.

[7] Hitchin N. J., On the Construction of Monopoles, *Comm. Math. Phys.* **89** (1983), pp. 145–190.

[8] Jaffe A. and Taubes C. H., *Vortices and Monopoles*, Birkhäuser, Boston, 1980.

[9] Murray M. K., Monopoles and Spectral Curves for Arbitrary Lie Groups, *Comm. Math. Phys.* **90** (1983), pp. 263–271.

[10] Nahm W., *All Self-Dual Multimonopoles for Arbitrary Gauge Groups*, preprint TH. 3172 – CERN, 1981.

[11] Prasad M. K. and Sommerfield C. M., Exact Solutions for the 't Hooft Monopole and the Julia–Zee Dyon, *Phys. Rev. Lett.* **35** (1975), pp. 760–762.

[12] Prasad M. K., Yang–Mills–Higgs Monopole Solutions of Arbitrary Topological Charge, *Comm. Math. Phys.* **80** (1981), pp. 137–149.
[13] Taubes C. H., *The Existence of a Non-Minimal Solution to the* SU(2) *Yang–Mills–Higgs Equations of* $\boldsymbol{R}^3$, preprint HUTMP 82/B118, 1982.
[14] Ward R. S., A Yang–Mills–Higgs Monopole of Charge 2, *Comm. Math. Phys.* **79** (1981), pp. 317–325.
[15] Yang C. N., Fibre Bundles and the Physics of the Magnetic Monopole. In: *The Chern Symposium 1979*, Springer-Verlag, New York.

ST. CATHERINE'S COLLEGE, OXFORD, ENGLAND

**Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa**

A. G. KHOVANSKIĬ

Fewnomials and Pfaff Manifolds

The ideology of fewnomials implies that real varieties, defined by "simple" (not too complicated) sets of equations, must have a simple topology. Of course, this is not always true. The fewnomial ideology, however, is helpful in finding a number of rigorous results.

The classical Bézout theorem states that the number of complex solutions of a set of k polynomial equations in k unknowns can be estimated in terms of their degrees (it equals the product of the degrees). This report is concerned with the real and the transcendental analogues of this theorem: for a wide class of real transcendental equations (including all real algebraic ones) the number of solutions of a set of k such equations in k real unknowns if finite and can be explicitly estimated in terms of the "complexity" of the equations. A more general result involves a construction of a class of transcendental real varieties resembling algebraic varieties.

These results provide new information about polynomial equations (see Sections 1 and 11) and level sets of elementary functions (see Sections 2 and 10).

1. Real fewnomials

The topology of geometric objects determined by algebraic equations (real algebraic curves, surfaces, singularities, etc.) gets more and more complex as the degree of the equation increase. As recently found complexity of the topology depends only on the number of monomials contained in the equations rather than on their degrees: the following Theorems 1 and 2 assess the complexity of the topology of geometrical objects in terms of the complexity of equations determining the object.

We begin with the following well-known

DESCARTES RULE. *The number of positive roots of a polynomial in a single real variable does not exceed the number of sign alternations in the sequence of its coefficients* (null coefficients are deleted from the sequence).

COROLLARY (the Descartes estimate). *The number of positive roots of a polynomial is less than the number of its terms.*

A. G. Kushnirenko proposed to call polynomials with a small number of terms *fewnomials*. The Descartes estimate shows that independently of the degree of a fewnomial (which may be as large as we wish) the number of its positive roots is small.

The following Theorems 1 and 2 (see [9]) generalize the Descartes estimate to the case of systems of polynomial equations in multi-dimensional real space.

Denote by q the number of monomials appearing with nonzero coefficients in at least one of the polynomials of the system.

THEOREM 1. *The number of non-degenerate solutions of a system of k polynomial equations in k positive real unknowns is less than* $2^{q(q-1)/2}(k+1)^q$.

THEOREM 2. *The sum of Betti numbers of a non-singular algebraic manifold defined in R^k by a non-degenerate system of polynomial equations is not greater than an explicitly expressed function of k and q. The number of connected components of a singular algebraic variety can also be estimated from above in terms of k and q.*

The known estimates of the sum of Betti numbers and of the number of connected components in Theorem 2, as well as the estimate of the number of roots in Theorem 1 contain an unpleasant factor of order $2^{q^2/2}$. Apparently, these estimates are far from being exact.

The arguments proving Theorems 1 and 2 are not only useful in algebra. Let us state a result related to the theory of elementary functions.

2. Level surfaces of elementary functions

We begin with definitions. Here is a list of *principal elementary functions*: the exponent, the logarithm, trigonometric functions (sin, cos, tan, cot) and their inverse functions. The function defined in a domain in R^n which can be represented as a composition of a finite number of algebraic functions and principal elementary functions is called *elementary*. An *elementary manifold* is the transversal intersection of non-singular level surfaces of several elementary functions. A map of degree $\leqslant m$ of an elementary manifold in R^n is the restriction to the manifold of such a map of R^n into R^k that all its components are polynomials of degree $\leqslant m$.

Choose a compact subset K in some k-dimensional elementary manifold.

THEOREM. *In any regular value in R^k of a map of degree m of a k-dimensional elementary manifold, the number of points in the inverse image contained in K is less than Cm^r. In this estimate the constant r depends only on the elementary manifold, while the constant C depends on the choice of the set K as well.*

This theorem may be applied to complex level surfaces of elementary functions in several complex variables and to the intersections of such surfaces. The theorem has various generalizations. For example, it remains valid if instead of level surfaces of elementary functions we consider level surfaces of functions which can be represented in quadratures.

COROLLARY. *The number of isolated intersection points of a compact arc of the graph of an elementary function in one variable with a plane algebraic curve of degree m is no greater than Cm^r. If for a compact arc of the graph of a certain function f there exists a sequence of algebraic curves of degrees m_n, whose number of intersection points with the arc increases faster than any power of the numbers m_n, then the function f is necessarily non-elementary.*

A similar situation is known in the theory of numbers: according to the Liouville theorem algebraic numbers have only "slow" approximations by rational numbers, so that if for some number there exists a. "rapid" approximation, then this number is necessarily transcendental-

Note that elementary functions have special properties also as multi valued functions of a complex variable [5].

3. Pfaff curves

There exists a wide class of real analytic manifolds whose properties are similar to those of algebraic manifolds. It is precisely this fact that the theorems of the previous section are based on.

Now we pass to the simplest variant of this theory. Consider a dynamic system on a plane given by a polynomial vector field. The trajectories of such a system may differ drastically from algebraic curves. There is no analogue of the Bézout theorem for such lines. Thus, for example, the trajectory of the system winding around a cycle, has a countable number of points in common with any straight line intersecting the cycle. The following constraint on the topology of the trajectory affects its properties and makes them similar to those of an algebraic curve.

DEFINITION. An oriented, smooth (possibly disconnected) curve in a plane is said to be a *separating solution of a dynamic system* if (a) the curve consists of one or several trajectories of the system (with natural orientation of trajectories); (b) the curve does not pass through the singular points of the system; (c) the curve is the boundary of a certain plane domain equipped with a natural boundary orientation.

Examples. (1) The cycle of a dynamic system is always its separating solution: It is oriented either as the boundary of the interior domain with respect to the cycle, or as the outer domain boundary. (2) A non-compact trajectory tending to infinity for $t\to+\infty$ and for $t\to-\infty$ is a separating solution. (3) A non-critical level line of a function $H(x_1, x_2) = c$, oriented as the boundary of the domain $H < c$, is a separating solution of the Hamiltonian system $\dot{x}_1 = -\partial H/\partial x_2$, $\dot{x}_2 = \partial H/\partial x_1$.

DEFINITION. A curve on a plane is called a *Pfaff curve of degree n* if there is an orientation of the curve for which the curve is a separating solution of a dynamic system, given by a vector field whose components are polynomials of degree n.

Smooth algebraic curves of degree $n+1$ are the Pfaff curves of degree n (see Example (3)). Thus the Pfaff curves can be viewed as generalizations of plane algebraic curves.

THEOREM (The analogue of the Bézout theorem for Pfaff curves). (1) *Restrictions of a polynomial of degree m to a Pfaff curve of degree n have at most $m(n+m)$ isolated roots.* (2) *Two Pfaff curves of degrees n and m have at most $(n+m)(2n+m)+n+1$ isolated intersection points.*

Let us consider the direct corollaries of this theorem.

COROLLARIES. (1) *All cycles of a dynamic system with polynomial field of degree 2 are convex.* (2) *Restrictions of a polynomial of degree m to a Pfaff curve of degree n have at most $(n+m-1)(2n+m-1)$ critical values on this curve.* (3) *A Pfaff curve of degree n has at most $n+1$ non-compact components and at most $(3n-1)(4n-1)$ inflection points.*

Of course, these estimates, both in the theorem and the corollaries are not the best possible (this is indicated, in particular, by the asymmetry with respect to n and m in the second statement of the theorem). However, these estimates are not so bad. Thus, for any m and any $n > 0$ it is easy to construct examples where the number of roots in the first statement of the theorem is not smaller than one third of the respective estimate. The estimate of the number of non-compact components is sharp and the estimates of the number of compact components and of the number of inflection points have the same order of growth for $n \to \infty$ as the sharp estimates for algebraic curves of degree $n+1$.

We now pass to the multidimensional case.

4. Separating solutions and Rolle's theorem

Let M be a smooth manifold (possibly disconnected, non-oriented and infinite-dimensional) and let α be a 1-form on it. Of great significance for the sequel is the following generalization of the separating solution of dynamic system on a plane.

DEFINITION. A submanifold of codimension one in M is said to be a *separating solution of the Pfaff equation* $\alpha = 0$ if (a) the restriction of the form α to the submanifold is identically zero; (b) the submanifold does not pass through the singular points of the equation (i.e., at each point of the submanifold the form α does not vanish on the tangent space); (c) the submanifold is a boundary of a domain in M and its coorientation defined by the form coincides with the coorientation of the domain boundary (i.e., on the vectors, applied at the submanifold points and outgoing from the domain, the form α is positive).

Example. The surface $H = c$ of a non-singular level of the function H is a separating solution of the equation $dH = 0$ (it bounds the domain $H < c$).

A *Pfaff hypersurface in* R^n is a separating solution of the equation $\alpha = 0$ where α is a 1-form in R^n with polynomial coefficients. An algebraic hypersurface is a Pfaff hypersurface (see the example). The Pfaff hypersurface resembles an algebraic one in many ways. Suppose β is the restriction of the 1-form with polynomials coefficients to the Pfaff hypersurface. A separating solution of the equation $\beta = 0$ on the Pfaff hypersurface also possesses properties similar to those of an algebraic manifold. This process may be continued. We obtain a wide class of manifolds resembling algebraic ones. The formal definition of this class is given in Sections 5, 7.

Here we will dwell on a certain property of separating solutions. For such solutions we have the following multidimensional variant of Rolle's theorem.

PROPOSITION. *Between two intersection points of a connected smooth curve with a separating solution of a Pfaff equation there is a point of contact, i.e., a point at which the tangent vector to the curve lies in hyperplane* $\alpha = 0$.

The proof is especially easy in the case where the curve intersects the separating solution transversally. In this case, at the neighbouring points of intersection, the values of the form α on the tangent vectors orienting the curve have different signs. Therefore, the form α vanishes at a certain intermediate point.

To demonstrate the significance of Rolle's theorem, we consider a simple transcendental generalization of the Descartes estimate.

PROPOSITION (Laguerre). *The number of real roots of a linear combination of exponents* $\sum_{i=1}^{q} \lambda_i \exp(a_i t)$ *is less than the number of exponents* q.

The Descartes estimate of the number of positive roots of a polynomial follows from the Laguerre proposition by substitution $x = \exp t$.

The proposition is proved by induction. Let us divide the linear combination by one of its exponents and differentiate the quotient. The derivative contains fewer exponents. According to Rolle's theorem, the number of zeroes of the function does not exceed the number of zeroes of the derivative plus 1.

Pfaff manifold theory is something of a multidimensional generalization of this simple argument (unidimensional generalization can be found in [4]).

5. Simple Pfaff manifolds

Let X be a real analytic manifold and A — a certain finitely generated ring of analytical functions on it.

DEFINITION. The following set of objects is called a *simple realization of the pair* (X, A): (a) an embedding π: $X \to R^n$ such that ring A coincides with the image of the polynomial ring under π^*; (b) a chain of embedded submanifolds in R^n, $X_0 \supset X_1 \supset \ldots \supset X_q$ in which every manifold is a hypersurface in the preceding one, the first manifold X_0 coincides with R^n and the last manifold X_q contains the image of X under the embedding π as one or several of its connected components; (c) a chain of polynomial 1-forms $\alpha_1, \ldots, \alpha_q$ in R^n such that each manifold X_i in the chain is a separating solution of the Pfaff equation $\alpha_i = 0$ on the preceding manifold X_{i-1}.

DEFINITION. A pair (X, A) is called a simple *Pfaff A-manifold* (briefly, A-manifold) if a simple realization for it exists.

DEFINITION. The *complexity of a simple realization of the pair* (X, A) is the set of degrees of all the polynomials which are the coefficients of all 1-forms α_i appearing in its simple realization. By the *degree of a function from ring* A at a simple realization we mean the minimal degree of the polynomial sent to this function by π^*.

In the following two examples, the pair (X, A) is defined simultaneously with its simple realization. The manifold X is defined as a transversal intersection of q non-singular hypersurfaces $f_i = 0$ in R^n, the ring A as the restriction to X of the polynomial ring and the i-th manifold in the chain as the intersection of the first i hypersurfaces (i.e. X_i is determined by the system $f_1 = \ldots = f_i = 0$).

Examples. (1) Consider algebraic hypersurfaces $f_i = 0$ (the functions f_i are polynomials). The chain of forms is $\alpha_i = df_i$. The complexity of realization is determined by the degrees of equations $f_i = 0$ defining the manifold X. (2) Take hypersurfaces defined by equations $f_i = y_i - \exp\langle a_i, t\rangle$ (here $R^n = R^k \times R^q$, $t \in R^k$, $a_i \in R^{k*}$ and y_i is the i-th component of the vector $y \in R^q$). For the chain of forms, we take the forms $\alpha_i = dy_i - y_i\langle a_i, dt\rangle$. The complexity of the realization is determined by the number q of exponents $\exp\langle a_i, t\rangle$ appearing in the definition of the manifold X.

THEOREM (Bézout theorem for simple Pfaff manifolds). *On a simple A-manifold of dimension k the number of non-degenerate solutions of a system $\varphi_1 = \ldots = \varphi_k = 0$, where $\varphi_i \in A$, is finite. The number of solutions is explicitly estimated by the complexity of any realization of the pair (X, A) and in terms of the degrees of the functions φ_i at the realization.*

Let us quote an example of an explicit estimate. Suppose that for a certain realization of the pair (X, A) the codimension of X in R^n equals q, that the degrees of all polynomial coefficients of the forms α_i do not exceed m, and the degrees of functions $\varphi_1, \ldots, \varphi_k$ are equal to $p_1, \ldots, p_k$.

ESTIMATE. *Under the above conditions, the number of solutions in our* *theorem does not exceed*

$$2^{q(q-1)/2}\left[p_1\cdot\ldots\cdot p_k\left(\sum(p_i-1)+mq-1\right)^q\right].$$

Indeed, our proof results in a more accurate but more awkward estimate (the difference is especially significant in the case where the coefficients of different forms a_i have different degrees or have higher degrees but smaller-volume Newton polyhedra). However, there is an unpleasant factor of the order $2^{q^2/2}$ inherent to our technique (improvements relate to the second factor which grows slower).

Let us consider a set of k equations $P_1 = \ldots = P_k = 0$ in k unknowns $(t_1, \ldots, t_k) = t$ in which P_j is a degree p_j polynomial in $k+q$ variables (t, y), where $y = (y_1, \ldots, y_q)$ and $y_i = \exp\langle a_i, t\rangle$.

THEOREM. *The number of non-degenerate real solutions of the system considered is finite and does not exceed*

$$2^{q(q-1)/2}p_1\cdot\ldots\cdot p_k\left(\sum p_i+1\right)^q.$$

To prove it, it suffices to apply the previous theorem to the manifold X from Example 2. Theorem 1 is a corollary of the formulated theorem (derived from it by the substitution of coordinates $x_i = \exp t_i$).

Let us now proceed to general Pfaff manifolds obtained by glueing simple manifolds.

6. Finiteness theorems

In the sequel we shall give definitions of a class of Pfaff real analytic manifolds, functions, forms and maps. A notion of a realization is introduced for each of these objects (one object has many different realizations). To each realization we assign a set of integers called its complexity.

THEOREM (analogue of the Bézout theorem). *The number of points in a Pfaff zero-dimensional manifold is finite and is explicitly estimated in terms of the complexity of any of its realizations.*

A Pfaff manifold is called *affine* if it holds at least one Morse–Pfaff function defining its proper map into the straight line R^1. The choice of such a function together with its realization is called the *realization of the affine manifold.*

FINITENESS THEOREM. *A Pfaff affine variety has the homotopy type of a finite cellular complex. The number of cells is explicitly bounded from above in terms of the complexity of any of its realizations.*

COROLLARY. *The sum of Betti numbers of an affine Pfaff manifold is finite and explicitly bounded in terms of the complexity of any of its realizations.*

Let us proceed to defining the category of Pfaff manifolds.

7. Pfaff manifolds, functions, forms, and maps

We call a ring of analytic functions on an analytic manifold a *base,* if: (a) the ring is finitely generated; (b) for any two different points of the manifold there is a function from the ring having different values at these points; (c) differentials of the ring function generate cotangent spaces in each point of the manifold.

A *Pfaff cover* of a manifold X with the base ring A is, by definition, a representation of X as a finite sum of open sets $X = \bigcup U_i$ together with the rings A_i of analytic functions in the domains U_i such that (a) the ring A_i contains all the functions that are restrictions of functions from the ring A to the domain U_i, (b) all pairs (U_i, A_i) are simple A_i-manifolds.

DEFINITION. A pair (X, A) is called a *Pfaff A-manifold* if it has a Pfaff cover. A function φ on the manifold X is called a *Pfaff A-function* if there is a Pfaff cover for which the restrictions φ_i of φ to the domains U_i lie within the rings A_i.

PROPOSITION. *Let B be any base ring consisting of A-functions on an A-manifold X. Then X is a B-manifold. Moreover, classes of A-functions and B-functions coincide.*

DEFINITION. A manifold with a function ring K is called a *Pfaff manifold* and functions of the ring K are called *Pfaff functions* if for a certain (and hence for any) base ring of functions $A \subset K$ the manifold is an A-manifold and the ring K coincides with the ring of A-functions.

The differential forms lying in the exterior algebra generated by the Pfaff functions and their differentials are referred to as *Pfaff forms.*

A mapping $\varphi\colon X \to Y$ of Pfaff manifolds with rings K_X and K_Y is called a *Pfaff map* if $\varphi^* K_Y \subseteq K_X$.

PROPOSITION. *Suppose that the Pfaff functions $\{f_i\}$ generate a base ring on the manifold Y. The map $\varphi\colon X \to Y$ is a Pfaff map if and only if the functions $\{\varphi^* f_i\}$ are Pfaff functions on X.*

8. Realizations and their complexity

By a *realization of a Pfaff manifold* with a ring K we mean a choice of a base ring $A \subseteq K$, a Pfaff cover $\{U_i, A_i\}$ and simple realizations of pairs U_i, A_i. The set of complexities of these simple realizations is called the *complexity of the realization.*

A *realization of the function* $\varphi \in K$ is such a realization of the manifold that the restriction φ_i of the function φ to each domain U_i is in the ring A_i. The *complexity of a realization of a function* φ is the complexity of the realization of the manifold together with the set of degrees of functions φ_i for the respective simple realizations of pairs U_i, A_i.

By a *realization of a Pfaff map* $\varphi\colon X \to Y$ we mean a choice of a base of Pfaff functions $\{f_i\}$ on Y together with the choice of realizations of functions $\{f_i\}$ on Y and functions $\{\varphi^* f_i\}$ on X.

A *realization of a Pfaff form* is a choice of its representations in terms of Pfaff functions and their differentials together with the choice of the functions' realizations.

Finally, by the *complexity of a realization* of a map or of a form we mean the set of complexities of the realizations of the functions involved.

9. Operations on Pfaff manifolds

PROPOSITION 1. *On a smooth real algebraic manifold (affine or projective) there exists only one function ring containing a ring of non-singular rational functions on the algebraic manifold and transforming it into a Pfaff manifold. The corresponding algebraic maps are Pfaff maps.*

PROPOSITION 2. *Let an analytic manifold be embedded in a Pfaff manifold. Then, there exists at most one function ring on the analytic manifold transforming it into a Pfaff manifold for which the embeddding is a Pfaff map.*

If the ring under the conditions of Proposition 2 does exist, then the manifold together with that ring is called a *Pfaff sub-manifold.*

In each of the following cases 1–4, the domain in a Pfaff manifold is a Pfaff submanifold (subdomain).

1. The domain consisting of one or several connected components of the manifold.

2. The domain defined by the inequality $f \neq 0$, where f is a Pfaff function.

3. The domain defined by the inequality $f > 0$, where f is a Pfaff function.

4. The domain being the complement to the zero set of a certain Pfaff form.

In each of the following cases 5–6 the submanifold is a Pfaff manifold.

5. The submanifold being an inverse image of a regular value under a Pfaff map.

6. The submamifold being a hypersurface and a separating solution of a Pfaff equation $\alpha = 0$ for a certain Pfaff 1-form α.

Here are two more operations on Pfaff manifolds.

7. The product of a finite number of Pfaff manifolds is a Pfaff manifold. To be exact, the product has a single function ring transforming it into a Pfaff manifold such that the projections onto the factors are Pfaff maps.

8. Let a manifold X, equipped with a function ring A, be covered by a finite number of domains, U_i, and let π_i: $U_i \to X$ be the embedding maps. If all the domains U_i are $(\pi_i^* A)$-manifolds, then X is an A-manifold.

For all cases 1–8 realization of all the manifolds constructed can be explicitly obtained from any realizations of the objects determining the construction. Its complexity is explicitly bounded from above in terms of the realization of the objects involved.

Note that affine and projective (!) real algebraic manifolds are affine Pfaff manifolds, and that operations 1–7 leave the manifolds inside the class of affine manifolds.

10. Properties and examples of Pfaff functions and Pfaff manifolds

1. Pfaff functions form a ring.

2. If a Pfaff function f does not vanish anywhere, then f^{-1} is a Pfaff function.

3. If w_1 and w_2 are Pfaff forms of a higher degree and w_2 does not vanish anywhere, then w_1/w_2 is a Pfaff function.

4. Let a domain in R^n be a Pfaff domain. Then the Pfaff functions in this domain form a differential ring (all partial derivatives of Pfaff functions are Pfaff functions).

5. Superpositions of Pfaff maps are Pfaff maps. In particular, the class of Pfaff functions is closed with respect to superposition.

6. If a vector function $y = (y_1, \ldots, y_k)$ satisfies a non-degenerate set of equations $F(x, y(x)) = 0$ where $F = F_1, \ldots, F_k$ are Pfaff functions, then $y_1, \ldots, y_k$ are Pfaff functions.

The most important property of the class of Pfaff functions is that it is closed with respect to the solution of Pfaff equations. Let us formulate this property more precisely.

Let M^{n+1} and M^n be $(n+1)$-dimensional and n-dimensional Pfaff manifolds, let $\pi\colon M^{n+1}\to M^n$ and $y\colon M^{n+1}\to R^1$ be a Pfaff map and a Pfaff function, α be a Pfaff 1-form on M^{n+1}, $\Gamma\subset M^{n+1}$ a separating solution of a Pfaff equation $\alpha = 0$ on M^{n+1} and let $\hat\pi$ be the restriction of the map π to Γ.

PROPOSITION. *If the projection $\hat\pi$ is a bijective bianalytic correspondence between Γ and M^n, then the function $y\circ\hat\pi^{-1}\colon M^n\to R^1$ is a Pfaff function on M^n.*

COROLLARY. *Suppose that the function $y(t)$, defined on a finite or infinite interval of a straight line, satisfies the differential equation $y' = F(t, y)$, where F is a Pfaff function in the plane or in its domain. Then y is a Pfaff function.*

It follows from the corollary that functions $\exp t$ and $\arctan t$ on the straight line, $\ln t$ and t^a on the ray $t>0$, $\arcsin t$ and $\arccos t$ on the interval $-1<t<1$ are Pfaff functions. The functions $\sin t$ and $\cos t$ are not Pfaff functions on the straight line as they have an infinite number of zeroes. However, they are Pfaff functions on any finite interval $a<t<b$. On the interval $0<t<\pi/2$ the functions sin and cos satisfy the equation $y' = \sqrt{1-y^2}$. The complexity of the minimal realization of these functions on the interval (a, b) is proportional to the integer part of the number $(b-a)/\pi$.

The collection of Pfaff functions on an algebraic variety is much wider than that of algebraic functions. Here are examples of Pfaff functions: $\exp\varphi$, $\arctan\varphi$, $\ln f$, f^a, $\arccos g$, $\arcsin g$, $\sin h$, $\cos h$ where φ, f, g, h are algebraic functions, and $f>0$, $-1<g<1$ and $a<h<b$. The polynomials in the above functions are again Pfaff functions; exponents (etc.) of these polynomials are again Pfaff functions, etc. (see [10]).

Non-singular level surfaces of a Pfaff function on algebraic manifold provide non-trivial examples of Pfaff manifolds. A more general example is given by the intersections of level surfaces of different functions. The finiteness theorem reveals that the sum of Betti numbers of such manifolds can be estimated by the complexity of realizations of functions determining them.

Let us return to the algebra.

11. Complex fewnomials

Complex roots of an elementary binomial equation $z^N-1=0$ of degree $N\to\infty$ are equidistributed with respect to arguments. The theorem formulated below shows that a similar phenomenon is observed for a fewnomial system of equations in k variables. The real fewnomial theorem (see Section 1) is one of the manifestations of this.

First, recall some definitions. The *support* of a polynomial $\sum C_\alpha z^\alpha$ depending on k complex variables is the set of degrees of monomials it contains, i.e., the finite set of points α in the integer lattice of space R^k for which the coefficients C_α are not equal to zero.

The *Newton polyhedron* of a polynomial is the convex hull of its support.

We shall denote a non-degenerate set of k polynomial equations in k complex unknowns by $P=0$. Denote by $T^k=(\varphi_1,\ldots,\varphi_k) \bmod 2\pi$ the torus arguments of space C^k (the j-th coordinate z_j of vector $z\in C^k$ is $z_j=|z_j|\exp i\varphi_j$). Let G be a domain in T^k. We are interested in the number $N(P,G)$ of solutions of the set $P=0$ for which all coordinates are non-zero, and their arguments lie in the domain G. In the case $G=T^k$ this number is determined by Bernstein's theorem (see [3], [6–8], [13]): it equals the mixed volume of the Newton polyhedra of the equations multiplied by $k!$. Let us denote by $S(P,G)$ the number from Bernstein's theorem multiplied by the ratio of the volume of G to the volume of T^k. For a certain number $\Pi(\Delta,\partial G)$ depending only on the domain G and the Newton polyhedra of the equations, the following theorem holds.

THEOREM ([11]). *There exists a function φ of k and q such that for any non-degenerate system $P=0$ of equations in k unknowns containing q monomials the following relation holds:*

$$|N(P,G)-S(P,G)|<\varphi(k,q)\,\Pi(\Delta,\partial G).$$

Let us present the definition of the number $\Pi(\Delta,\partial G)$. Let Δ be a domain in R^k determined by the set of inequalities $\{|\langle\alpha,\varphi\rangle|<\pi/2\}$ corresponding to a set of integer vectors α lying in the unions of the supports of the equations. The number $\Pi(\Delta,\partial G)$ is the least number of parallel translates of the domain Δ needed to cover the boundary of G. As corollaries we obtain two old theorems: (1) Bernstein's theorem is obtained for $G=T^k$, since in this case $\Pi(\Delta,\partial G)=0$; (2) Theorem 1 (Section 1) on real fewnomials is obtained when G contracts to the point $0\in T^k$; in this case $\Pi(\Delta,\partial G)=1$, $S(P,G)\to 0$.

For sets of equations with large Newton polyhedra, the number $S(P, Q)$ exceeds, in order of magnitude, the number $\Pi(\Delta, \partial G)$ (see [11]). Therefore, the theorem suggests a uniform distribution of roots of a fewnomial equation with respect to arguments.

A few words about proof of the theorem. First, it is shown that the average number $N(Q, G)$ coincides with the number $S(P, G)$. Averaging is performed for all systems $Q = 0$ whose equations have the same supports as the equations of the original set $P = 0$. This part of the proof is considerably clarified by the Atiyah paper on momentum mappings under almost periodic simplicial actions of the torus ([1], [2]). Then it is shown that numbers $N(Q, G)$ for different systems do not differ much. This part of the proof is based on the theorem from Section 5.

12. Some problems

1. The problem of A. G. Kushnirenko: *Find an exact estimate of the number of real roots of a fewnomial system. Give an example of a fewnomial system with the largest possible number of roots.* (The investigation of fewnomials originated with Kushnirenko's problem. The first result is due K. A. Sevastianov: he estimated the number of zeroes of a fewnomial on an algebraic curve. The first multidimensional result is the theorem on real fewnomials in Section 1.)

2. According to the Descartes rule (see Section 1) a polynomial with a large number of terms has few positive roots if the sign in its coefficient sequence rarely changes. *Find a multidimensional analogue of the Descartes rule* (compare [12]).

3. *Is there any analogue of the Seidenberg–Tarski theorem* [14] *for Pfaff manifolds?* Probably there exists such an analogue for a narrow class of varieties including the class of algebraic varieties. (Added in proof: see [21].)

4. *How can one extend the class of Pfaff manifolds and yet retain the finiteness theorems?* (Added in proof: see [15].)

5. Here is a more specific problem. Let ω be a 1-form in the plane whose coefficients are polynomials of the n-th degree. Cosider the integral J of the form ω with respect to a compact component of the level line $H = c$ of a polynomial H of the $(n+1)$-th degree. When the parameter c ranges over intervals on which the integration curve is not restructured, the integral is an analytic function of the parameter. The problem (of V. I. Arnold) is: *to estimate the number of the isolated integral zeroes as a function of the parameter in this interval.* The Arnold problem is a linearization in the neighbourhood of Hamiltonian systems of Hilbert's sixteenth

problem about the number of limiting cycles of polynomial dynamical systems in the plane. (Equation $J(c) = 0$ is the linearization of the cycle birth condition from the level line $H = c$ of the Hamiltonian H under a perturbation of the Hamiltonian system by the vector field $\varepsilon j(\omega)$, where ε is a small number and j is the isomorphism between cotangent and tangent spaces induced by the standard simplicial structure in the plane.)

(Added in proof: This problem was studied in [16], [17], the latter paper considerably advances its solution.)

References

[1] Atiyah M. F., Convexity and Commuting Hamiltonians, *Bull. London Math. Soc.* **14** (1) (1982), pp. 1–15.

[2] Atiyah M. F., *Angular Momentum, Convex Polyhedra and Algebraic Geometry*, talk at Edinburg, Math. Soc., 4 February, 1983.

[3] Bernstein D. N., The Number of Roots of a System of Equations, *Functional Analysis* **9** (3) (1975), pp. 1–4.

[4] Gelfond O. A. and Khovanskiĭ A. G., On Real Liouville Functions, *Functional Analysis* **14** (2) (1980), pp. 52–53.

[5] Khovanskiĭ A. G., On Representation of Functions in Quadratures, *Uspekhi Mat. Nauk* **26** (4) (1971), pp. 251–252.

[6] Khovanskiĭ A. G., Newton Polyhedra and Torus Varieties, *Functional Analysis* **11** (4) (1977), pp. 56–64.

[7] Khovanskiĭ A. G., Newton Polyhedra and Complete Intersections, *Functional Analysis* **12** (1) (1978), pp. 51–61.

[8] Khovanskiĭ A. G., Geometry of Convex Bodies and Algebraic Geometry, *Uspekhi Mat. Nauk* **34** (4) (1979), pp. 160–161.

[9] Khovanskiĭ A. G., On a Class of Systems of Transcendental Equations, *Reports of the USSR Academy of Sciences* **255** (4) (1980), pp. 804–807.

[10] Khovanskiĭ A., *Théorème de Bézout pour les fonctions de Liouville*, Septembre 1981, I.H.E.S./M/81/45 (France).

[11] Khovanskiĭ A., Sur les racines complexes de systemes d'équations algèbriques, *C. R. Acad. Sc. Paris* **292** (1) (1981), pp. 937–940.

[12] Khovanskiĭ A. G., Newton Polyhedra and the Vector Field Index, *Uspekhi Mat. Nauk* **36** (4) (1981), p. 234.

[13] Kouchnirenko A. G., Polyèdres de Newton et nombres de Milnor, *Invent. Math.* **32** (1) (1976), pp. 1–32.

[14] Tarski A., *A Decision Method for Elementary Algebra and Geometry*, Berkeley, 1951.

Added in proof:

[15] Khovanskiĭ A. G., Real Analytic Manifolds with Finiteness Properties and Complex Abelian Integrals, *Functional Analysis* **14** (2) (1984), pp. 40–50.

[16] Petrov, G. S., On the Number of Zeros of Elliptic Integrals, *Functional Analysis* **14** (2) (1984), pp. 73–74.

[17] Varchenko A. N., A Uniform Estimate from Above of the Number of Zeros of Real Abelian Integral, Depending on a Parameter and Cycles of a Polynomial Dynamic System, *Functional Analysis* **14** (2) (1894), pp. 14–25.
[18] Khovanskiĭ A. G., Cycles of Dynamic Systems on a Plane and the Rolle Theorem, *Siberian Math. J.* **25** (3) (1984), pp. 198–203.
[19] Bertrand D., Sur les lemmes d, zeros, *J. de Saint Etienne* (to appear).
[20] Van den Dries L., Analytic Hardy Fields and Exponential Curves in the Real Plane, *American J. of Math.* (to appear).
[21] Van den Dries L., Remarks on Tarski's Problem Concerning ($\boldsymbol{R}$, +, ·, exp), *Proceedings Peano Conference, Florence (1983)* (to appear).
[22] Risler J. J., Additive Complexity and Zeros of Real Polynomials, *Siam J. Comput.* (to appear).
[23] Richardson D., Roots of Real Exponential Functions, *J. London Math. Soc.* **28** (1) (1983), pp. 46–56.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

WERNER MÜLLER

Spectral Geometry and Non-Compact Riemannian Manifolds

1. Introduction

Let X be a smooth oriented Riemannian manifold and let $\Delta_p = dd^* + d^* d$ denote the Laplacian acting on the space of p-forms. Δ_p is one of the most useful differential operators in geometry. But many of the applications require the assumption that X is compact or even closed. In this report we want to survey some results concerning the theory of the Laplace operator on certain non-compact Riemannian manifolds. Spectral analysis on complete Riemannian manifolds is an active research topic. However, we cannot discuss all these results here. The reader interested in a more detailed treatment should consult the articles cited at the end. Our point of view is that the study of the Laplacian, in particular of its spectrum, on a certain natural class of non-compact Riemannian manifolds might have some useful applications. Consider a complex projective algebraic variety $M \subset P^N_{\mathbf{C}}$ and let $X = M - \Sigma$ be the complement of its singular set, with the metric induced by the inclusion $X \subset P^N_{\mathbf{C}}$. Analysis on this space would be very interesting. On the other hand, one can consider manifolds which generalize locally symmetric spaces. For example, if we change the metric of $X \subset P^N_{\mathbf{C}}$ near Σ as in [26] to get a complete Riemannian manifold, then this should be a manifold of this type. In both cases, the simplest examples are Riemannian spaces with cone-like singularities and Riemannian manifolds with cusps. The cone-like case has been studied in detail by Jeff Cheeger. Manifolds with cusps have been considered by the author. We discuss both cases because, on the one hand, they are closely related, but on the other hand they show up essential differences.

2. Singular Riemannian spaces

In recent works [5], [6], [7] Jeff Cheeger has extended the theory of the Laplace operator on compact Riemannian manifolds to certain Riemannian spaces with singularities. The simplest example is that of a metric cone. If N^m is a Riemannian manifold, then the metric cone $C(N^m)$, on N^m, is the space $\mathbf{R}^+ \times N^m$ with the metric $dr^2 + r^2 g$. Here g is the metric on N. A space X^{m+1} is called a Riemannian space with an isolated conical singularity if there exists $p \in X$ such that $X - \{p\}$ is a smooth Riemannian manifold and p has a neighbourhood U such that $U - \{p\}$ is isometric to $C_{0,u}(N^m) = \{(r, x) \in C(N^m) \mid 0 < r < u\}$, for some u and N^m. By definition, analysis on X^{m+1} is analysis on the incomplete Riemannian manifold $X - \{p\}$. Local analysis outside the singularity is as in the nonsingular case. Global analysis on X can be reduced to global analysis on the cone $C(N^m)$. By using the method of separation of variables, global analysis on the cone is equivalent to the functional calculus for the Laplacian on the cross-section. Consider the Laplacian Δ_p restricted to the space of p-forms Λ_0^p with compact supports. If $p \neq m/2$ or $2p = m$ and $H^p(N; \mathbf{R}) = 0$, then there exists a natural self-adjoint extension $\bar{\Delta}_p$ of Δ_p to $L^2 \Lambda^p$. If $m = 2p$ and $H^p(N^{2p}; \mathbf{R}) \neq 0$, one has to choose ideal boundary conditions (cf. [5]). The Green operator is compact. Thus $\bar{\Delta}_p$ has a pure point spectrum consisting of eigenvalues of finite multiplicity: $0 \leqslant \lambda_0 \leqslant \lambda_1 \leqslant \ldots \to \infty$. The heat equation $\left(\frac{\partial}{\partial t} + \Delta_i\right) u = 0$, with $u|_{t=0} = u_0$, has a fundamental solution $E_i(z, z', t)$. The trace of the heat operator is given by $\int_X \operatorname{tr} E_i(z, z, t)$. It has an asymptotic expansion, as $t \to 0$, and it is of the form

$$\sum_{i}^{a_{j/2}} t^{-(m+1)/2+j/2} + b_i \log t + \Psi_i(N).$$

The coefficient $\Psi_i(N)$ is a non-local spectral invariant of the cross-section N. Moreover, in the non-smooth case an extra term $\log t$ appears. The role of the De Rham cohomology is now played by the L^2-cohomology spaces $\hat{H}^i_{(2)}(X)$. They differ from the usual L^2-cohomology $H^i_{(2)}(X)$ at most for $i = k, k+1$, where $\dim N = 2k$. This is due to the ideal boundary conditions one has to choose if $H^k(N^{2k}; \mathbf{R}) \neq 0$. Let $\mathscr{H}^i_{(2)} = \ker \bar{\Delta}_i$ be the space of L^2-harmonic i-forms. Then one has the Hodge Theorem: $\mathscr{H}^i_{(2)}(X) = \hat{H}^i_{(2)}(X)$. The space $\hat{H}^*_{(2)}(X)$ does not give the topological cohomology of X. However, $\hat{H}^*_{(2)}(X)$ is dual to the middle intersection homology $IH_*(X)$ introduced by Goresky and Mac Pherson (cf. [8]).

Suppose that $X = M^{m+1} \cup C_{0,1}(N^m)$. If $m = 2k-1$ then

$$\hat{H}^i_{(2)}(X) = \begin{cases} H^i(M), & i < k, \\ \mathrm{Im}\big(H^i(M, N) \to H^i(M)\big), & i = k, \\ H^i(M, N), & i > k. \end{cases}$$

For $m = 2k$ one has a corresponding description. An application of the heat equation method to the signature complex gives

$$\mathrm{sign}(X) = \int_X L(p) + \eta_N(0),$$

where $L(p)$ is the Hirzebruch polynomial in the Pontryagin forms of X and $\eta(0)$ is the Eta-invariant of N introduced by Atiyah, Patodi and Singer in [2]. Thus, Cheeger recovers the result of Atiyah, Patodi and Singer. In the same way one gets a Gauss–Bonnet formula. By inductive arguments J. Cheeger extends his results to n-dimensional pseudomanifolds. Recall that an n-dimensional pseudomanifold X^n is a finite simplicial complex such that every point p is contained in a closed n-simplex and every $(n-1)$-simplex is the face of at most two n-simplices. The metric g on X^n is such that $X^n - \Sigma^{n-2}$ is flat in the sense that every n-simplex is isometric to the interior of some linear n-simplex. Cheeger considers also more general metrics which are only piecewise smoothly quasi-isometric to g (cf. [7]). Thus, he can study the Laplacian on simplicial manifolds. Under certain topological conditions the L^2-cohomology $H^i_{(2)}(X) \overset{\mathrm{def}}{=} H^i_{(2)}(X - \Sigma^{n-2})$ coincides with the space $\mathscr{H}^i_{(2)}(X) = \{h \in L^2 |\ dh = \delta h = 0\}$. Moreover, $H^*_{(2)}$ is naturally isomorphic to the dual of the middle intersection homology $IH_*(X)$. The heat operator $\exp(-t\Delta_i)$ is again of trace class and its trace has an asymptotic expansion as $t \to 0$. This expansion contains extra negative half powers of t but no positive power of t. Thus, this is quite different from the smooth case. Using this asymptotic expansion one gets generalizations of curvature invariants of smooth manifolds to piecewise linear manifolds. The Lipschitz–Killing curvatures are of particular interest here. In both the smooth and the p. l. cases it turns out that the Lipschitz–Killing curvatures are given by (the same) specific linear combinations of the coefficients in the asymptotic heat expansion on p-forms. Let X^n be an n-dimensional pseudomanifold. Let σ^j be a simplex of X and let $\sigma^i \subset \sigma^j$ be any of its faces. Let $[\sigma^i, \sigma^j]$ denote the normalized dihedral angle at the face σ^i of σ^j. This is the angular measure of the unit normals to σ^i at $p \in \sigma^i$ which point into σ^j. We consider the scalar curvature R_1, which is one extreme case. To define the p. l. expression we introduce the the angle defect

$$\delta(\sigma^{n-2}) = 1 - \sum_{\sigma^{n-2} \subset \sigma^n} [\sigma^{n-2}, \sigma^n].$$

Then R_1 is given by

$$R_1 = \sum_{\sigma^{n-2}} \delta(\sigma^{n-2}) \operatorname{vol}(\sigma^{n-2}), \tag{1}$$

where $\operatorname{vol}(\sigma^{n-2})$ is the volume of σ^{n-2}. The corresponding formulas exist for the other Lipschitz–Killing curvatures. The expression (1) has been proposed by T. Regge [22] as the analogue of the total scalar curvature. Let X_i^n be a suitable sequence of piecewise flat approximations of a smooth Riemannian space M^n. Then the conjecture that the expression (1) converges to the total scalar curvature of M^n seems to be a very natural one. It turns out that the corresponding statement is true for all Lipschitz–Killing curvatures. This has been proved in joint work of J. Cheeger, W. Müller and R. Schrader [9], [10]. Using his method, J. Cheeger can derive a local formula for the signature of a pseudomanifold. Let X^{4l} be a closed oriented Riemannian pseudomanifold with piecewise flat metric which satisfies an additional condition reflecting that X has negligible boundary (cf. [6]). Then

$$\operatorname{sign}(X^{4l}) = \sum_{\sigma^0} \eta\big(L(\sigma^0)\big),$$

where $\eta\big(L(\sigma^0)\big)$ denotes the Eta-invariant of the link of σ^0. Another very interesting aspect of this is a local formula for the L-classes of p. l. manifolds. Let X^n be a closed, oriented pseudomanifold with piecewise flat metric. For each k, define an $n-4k$ chain by

$$c_{n-4k} = \sum_{\sigma^{n-4k}} \eta\big(L(\sigma^{n-4k})\big)\, \sigma^{n-4k}.$$

Then the chains c_{n-4k} are cycles whose homology class L_{n-4k} depends only on the p. l. structure.

In this report we could only touch some of the highly interesting results proved by Jeff Cheeger.

3. Riemannian manifolds with cusps

Let S be a Riemannian symmetric space of noncompact type and let G be the group of motions of S. Let Γ be a discrete torsion free subgroup of G such that $\Gamma\backslash S$ has finite volume. A very important problem is to study the spectrum of the Laplacian on $\Gamma\backslash S$ (or, more generally, of the algebra of G-invariant differential operators $D(S)$). If $\Gamma\backslash S$ is noncompact one has, besides the discrete spectrum, one or more continuous spectra which make the situation much more complicated than in the compact case. The whole subject originated with Selberg's famous paper [23]. Our point of view is that only the metric structure near infinity of the locally symmetric space $\Gamma\backslash S$ is important to get all results about the spectral resolution of the Laplacian. Thus we give up the assumption that the manifold is locally symmetric and we consider complete Riemannian manifolds of finite volume with additional conditions near infinity reflecting the presence of "cusps". We think that this approach can be even helpful to solving problems in the locally symmetric case. A suitable class of such manifolds should cover both locally symmetric spaces of finite volume and the examples discussed in the introduction. The simplest case is what we call a manifold with cusps. This conception generalizes locally symmetric spaces of finite volume and with constant negative curvature. By a slight modification one can cover all **R**-rank one locally symmetric spaces of finite volume (cf. [18], [19]).

A manifold with cusps is a Riemannian manifold X which has a decomposition into a compact manifold X_0 with boundary ∂X_0 and a finite number of ends X_i, $i = 1, \ldots, m$, called cusps. Each X_i is a manifold with boundary $\partial X_i = Y_i$ such that $X_0 \cap X_i = Y_i$, $i \geqslant 1$. Moreover, the X_i's, $i \geqslant 1$, are Riemannian warped products. This means that each X_i is diffeomorphic to $\mathbf{R}^+ \times Y_i$, where Y_i is a closed Riemannian manifold and the metric ds^2 on X_i may be written as $ds^2 = dr^2 + e^{-2r} g_i$. Here g_i denotes the metric tensor of Y_i. Let Γ be a discrete torsion free subgroup of $\mathrm{SL}(2, \mathbf{R})$ with finite covolume and let $\mathbf{H}$ be the upper half-plane with the Poincaré metric. $\Gamma\backslash\mathbf{H}$ is an example of a manifold with cusps. Analysis on manifolds with cusps is divided into (1) Analysis on the cusp, (2) Analysis on the compact piece X_0. By using the method of separation of variables one can reduce global analysis on the cusp to the functional calculus for the Laplacian on the cross-section. Since X is complete, the Laplacian Δ_p acting on the space of compactly supported p-forms has a unique self-adjoint extension $\bar{\Delta}_p$ to $L^2 \Lambda^p$. One essential difference from the cone-like case is that the Laplacian $\bar{\Delta}_p$ has now a continuous spectrum of finite multiplicity. We discuss first the Laplacian $\bar{\Delta}_0$ acting on functions. The heat equation $(\partial/\partial t + \Delta_0)u = 0$ with $u|_t = u_0$ has a "good" fundamental solution $E(z, z', t)$ ("good" in the sense of [19]). Via the Laplace transform one gets an expression for the resolvent kernel which is valid in a certain half-plane $\mathrm{Re}(\lambda) < C$. Then we apply methods of perturbation theory to study the resolvent in a neighbourhood of the spectrum. It turns out that almost all results about the spectral resolution of the Laplacian on locally symmetric spaces of $\mathbf{R}$-rank one can be extended to Riemannian manifolds with cusps. Another method has been developed by Colin de Verdiere [11], [12]. He has extended the approach of Lax and Phillips [16]. The spectrum of Δ_0 consists of an absolute continuous spectrum and a point spectrum. The absolute continuous spectrum is the interval $[N^2/4, \infty)$, $N = \dim X - 1$, with multiplicity equal to the number of cusps of X. The point spectrum consists of eigenvalues of finite multiplicity, with infinity as the only possible accumulation point. Since Δ_0 has a continuous spectrum, the heat operator $\exp(-t\Delta_0)$ is not of trace class and the integral of the heat kernel $E(z, z, t)$ over X does not exist. For simplicity assume that X has a single cusp $X_1 = \mathbf{R}^+ \times Y$. One can modify the heat kernel by subtracting the constant term $\int_Y E((r, y), z', t)\,dy$. Let $E_0(z, z', t)$ be the resulting kernel. Then $\int_X E_0(z, z, t)\,dz$ exists. This integral has an asymptotic expansion, as $t \to 0$, of the form $\sum a_{j/2} t^{-n/2+j/2} + b(4\pi t)^{-1/2}\log t + (4\pi t)^{-1/2}\frac{1}{2}\big(\zeta'(0) - (\zeta(0)-1)\gamma\big) + O(1)$. Here $\zeta(s)$ is the zeta function of the Laplacian on Y, γ is Euler's constant and the $a_{j/2}$ are locally computable. This expansion can be extended to an arbitrary order. Thus, there appears a term $\zeta'(0)$, which is a non local spectral invariant. There is also a logarithmic term which is absent in the compact case. Let Δ be the restriction of Δ_0 to the subspace $L^2_d(X) \subset L^2(X)$ spanned by the eigenfunctions. Then $e^{-t\Delta}$ is trace class and $\operatorname{Tr}(e^{-t\Delta}) = \int_X E_0(z, z, t)\,dz + C(t)$, where $C(t)$ is the contribution of the continuous spectrum to the trace formula. Very little is known about the function $C(t)$. For example, the existence of an asymptotic expansion as $t \to 0$ is in general unknown. Even for $X = \Gamma\backslash\mathbf{H}$, $\Gamma \subset \mathrm{SL}(2, \mathbf{R})$ discrete and of finite covolume, one does not know much about $C(t)$. Only for very special subgroups of $\mathrm{SL}(2, \mathbf{R})$ ($\Gamma_0(N)$, $\Gamma(N)$ for example) one has sufficiently good results. This is related to the fact that the behaviour of the discrete spectrum of Δ_0 is completely different from the compact case. Since the continuous spectrum consists of the interval $[N^2/2, \infty)$, almost all eigenvalues are embedded into the continuous spectrum. But embedded eigenvalues are very unstable. Thus the existence of infinitely many eigenvalues can be considered as an exceptional case. More precisely, let us fix a metric g_0 on X and consider conformal perturbations of g_0 on a compact subset $K \subset X$ with non-empty interior. Then the set $\mathscr{M}$ of all $f \in C_0^\infty(K)$ for which the perturbed Laplacian Δ_f has no eigenvalues $\lambda \geqslant N^2/4$ is a residual set in the sense of Baire. This result has been proved by Colin de Verdiere [12] in the 2-dimensional case. His method can easily be extend ed to the higher dimensional case [20]. This result should be seen in-contrast with the Roelke conjecture (cf. [25]) which claims that the Laplacian on $\Gamma\backslash\mathbf{H}$, Γ as above, always has infinitely many eigenvalues. Thus, at least generically, there exists an asymptotic expansion of $C(t)$ as $t\to 0$. In the compact case one has Weyl's formula for the asymptotic number of eigenvalues. For manifolds with cusps one has an asymptotic upper bound for the maximal possible number of eigenvalues. For $\lambda > 0$ let $N(\lambda)$ be the number of eigenvalues of Δ_0 (counted with multiplicity) which are less than λ. Then $N(\lambda)$ is finite for each $\lambda > 0$ and

$$\lim_{\lambda\to\infty} \frac{N(\lambda)}{\lambda^{n/2}} \leqslant (4\pi)^{-n/2} \frac{\mathrm{Vol}(X)}{\Gamma(n/2+1)}, \tag{2}$$

where $n = \dim X$. In the 2-dimensional case this has been proved by Colin de Verdiere [12], by using the Lax–Phillips approach. For locally symmetric spaces of $\mathbf{R}$-rank one, (2) was proved by Donnelly [13]. He then extended his own results to locally symmetric spaces of arbitrary rank, but $N(\lambda)$ had to be replaced by the number of such eigenvalues that correspond to cuspidal eigenfunctions [14]. For Riemannian manifolds with cusps one can use the trace formula combined with Neumann bracketing to prove (2) (cf. [20]). Everything that we have described for the Laplacian on functions can be extended to the Laplacian on p-forms. Assume that X has a single cusp $X_1 = \mathbf{R}^+ \times Y$. Every $\Phi \in \mathscr{H}^p(Y)$ defines a p-form $E(\Phi, z, s)$ and a $(p+1)$-form $E\,(dr\wedge\Phi, z, s)$ on X, depending on $s \in \mathbf{C}$. These forms are generalized eigenforms of the corresponding Laplacian. The continuous spectrum of Δ_p is non-empty iff $\mathscr{H}^p(Y)\oplus \oplus\mathscr{H}^{p-1}(Y) \neq 0$. If $\mathscr{H}^p(Y) = \mathscr{H}^{p-1}(Y) = 0$ then $\exp(-t\Delta_p)$ is of trace class and its trace is given by $\int_X \mathrm{tr}\, E_p(z, z, t)$ where $E_p(z, z', t)$ is the heat kernel on p-forms. In the asymptotic expansion again there appear logarithmic terms and non-local spectral invariants of the cross-section Y. Let $\mathscr{H}^p_{(2)}(X)$ be the space of square integrable harmonic p-forms and let $H^p_{(2)} = \ker d_p/\mathrm{range}\, d_{p-1}$ be the L^2-cohomology (cf. [27]). By Kodaira we have an injection $j\colon \mathscr{H}^p_{(2)}\to H^p_{(2)}$. It turns out that j is an isomorphism if $p \neq N/2+1$, $N = \dim X - 1$. Moreover, if $N = 2k$ then $H^{k+1}_{(2)}(X) = \mathscr{H}^{k+1}_{(2)}(X)$ iff $\mathscr{H}^k(Y) = 0$. If $\mathscr{H}^k(Y) \neq 0$ then $H^{k+1}_{(2)}(X)$ is infinite-dimensional. $\mathscr{H}^*_{(2)}(X)$ is again dual to the middle intersection homology $IH_*(\bar{X})$ of the one point compactification of X. For example, let $N = 2k$. Then

$$\mathscr{H}^p_{(2)}(X) \cong \begin{cases} H^p(X), & p < k, \\ \mathrm{Im}\big(H^p_c(X)\to H^p(X)\big), & p = k, \\ H^p_c(X), & p > k. \end{cases}$$

Here $H_c^p(X)$ is the cohomology with compact supports. Let $\dim X = 4k$. If we apply our method to the signature complex then we obtain

$$\operatorname{sign}(X) = \int_X L(p) - \eta(0),$$

where $L(p)$ is the Hirzebruch polynomial in the Pontryagin forms of X and $\eta(0)$ is the Eta-invariant of the cross-section Y. Thus, we recover the result of Atiyah, Patodi and Singer [2]. The role of the non-local boundary conditions on X_0 in their work corresponds to the L^2-condition on X in our case. The heat equation method can also be applied to the calculation of the Euler characteristic. The result is $\chi(X) = \int_X P(\Omega)$. Here $P(\Omega)$ denotes the Chern–Gauss–Bonnet form of X. Of course, this can also be proved by different methods.

It would be very important to extend the results of [18], [19] to the case where $X = \Gamma \backslash G/K$, Γ a rank one lattice in G. At the moment we do not know how to do this. What we can do is to study Riemannian manifolds which outside a compact set are isometric to the complement of a compact set in $\Gamma \backslash G/K$. Thus, we keep the locally symmetric structure near infinity fixed, but we can alter the metric on a compact set and even change the topological type of the manifold arbitrarily. We can do spectral analysis on such spaces, because global analysis on the "cusps" is reduced to harmonic. One application of this is a proof of the Hirzebruch conjecture [15, p. 230]. Investigating Hilbert modular surfaces, Hirzebruch discovered a very interesting relation between the signature defect associated to a cusp of a Hilbert modular surface and the value at $s = 1$ of certain L-series. Based on this result, Hirzebruch conjectured that for higher-dimensional Hilbert modular varieties the signature defects of cusps should still be given by values at $s = 1$ of corresponding L-series. The attempt to prove the Hirzebruch conjecture was one of the main motivations for the work of Atiyah, Patodi and Singer on spectral asymmetry [2]. The idea is to apply the results of [2] to the boundary of a neighbourhood of a cusp. A proof of the Hirzebruch conjecture along these lines has been developed by Atiyah, Donnelly and Singer [1].

Our proof of the Hirzebruch conjecture is a consequence of a certain L^2-index theorem for manifolds as above. For example, consider the Hilbert modular group. Let $F/\mathbf{Q}$ be a totally real number field of degree n and class number one. Let $\mathcal{O}_F$ be the ring of integers and consider the Hilbert modular group $\Gamma = \mathrm{SL}(2, \mathcal{O}_F)$. $\Gamma \backslash \mathbf{H}^n$ has a single cusp. Let $M = \mathcal{O}_F$ and $V = \mathcal{O}_F^{*2}$. The cusp is of type (M, V). Let $\Lambda^*(\Gamma \backslash \mathbf{H}^n)$ be the space of Γ-invariant differential forms on $\mathbf{H}^n$ and consider the signature operator $D = d+d^*$: $\Lambda_+^* \to \Lambda_-^*$. It has a well defined L^2-index which is equal to $\operatorname{sign}(\Gamma\backslash\mathbf{H}^n)$. Using the heat equation method combined with Selberg's trace formula, one can compute the L^2-index in a different way. Let $z_1, \ldots, z_m \in \mathbf{H}^n$ be points which represent the quotient singularities of $\Gamma\backslash\mathbf{H}^n$ and let $\delta(z_j)$ be the cotangent sum associated with it [15, § 3]. Then

$$\operatorname{sign}(\Gamma\backslash\mathbf{H}^n) = \sum_{j=1}^{m} \delta(z_j) + \frac{i^n}{\pi^n} d(M) L(M, V, 1), \tag{3}$$

where $d(M) = (D_{F/\mathbf{Q}})^{1/2}$ and $L(M, V, s) = \sum_{\mu\in(M-0)/V} \operatorname{sign} N(\mu)/|N(\mu)|^s$. If we compare (3) with Hirzebruch's formula for $\operatorname{sign}(\Gamma\backslash\mathbf{H}^n)$ in [15], we find that the signature defect $\delta(\infty)$ of the cusp ∞ is equal to $\frac{i^n}{\pi^n} d(M) L(M, V, 1)$. This is the Hirzebruch conjecture in this case.

References

[1] Atiyah M. F., Donnelly H., and Singer I.M. , Eta Invariants, Signature Defects, of Cusps, and Values of L-Functions, *Annn. of Math.* **118** (1983).

[2] Atiyah M. F., Patodi V. K., and Singer, I. M., Spectral Asymmetry and Riemannian Geometry I, *Math. Proc. Camb. Phil. Soc.* **77** (1975), pp. 43–69.

[3] Bérard P. H., and Berger M., *Le spectre d'une varieté Riemannienne en 1981*, preprint.

[4] Cheeger J., Analytic Torsion and the Heat Equation, *Annals of Math.* **109** (1979), pp. 259–322.

[5] Cheeger J., On the Spectral Geometry of Spaces with Cone-Like Singularities, *Proc. Nat. Acad. Sci. USA* **76**, No. 5 (1979), pp. 2103–2106.

[6] Cheeger J., *Spectral Geometry of Singular Riemannian Spaces*, preprint, 1982.

[7] Cheeger J., On the Hodge Theory of Riemannian Pseudomanifolds, *Proc. Symp. Pure Math.* **36** (1980), pp. 91–145.

[8] Cheeger J., Goresky M. and MacPherson R., L^2-Cohomology and Intersection Homology of Algebraic Varieties, *Ann. of Math. Studies* **102**, Princeton Univ. Press, 1982, pp. 303–340.

[9] Cheeger J., Müller W., and Schrader R., Lattice Gravity or Riemannian Structures on Piecewise Linear Spaces, *Proc. of Heisenberg Symp., München 1981* (to appear).

[10] Cheeger J., Müller W., and Schrader R., *Comm. Math. Physics* **92** (1984).

[11] Colin de Verdiere Y., Une nouvelle preuve du prolongement meromorph des series d'Eisenstein, *Comptes-Rendus de l'Acad. des Sciences de Paris* **293** (1981), pp. 361–363.

[12] Colin de Verdiere Y., Pseudo-Laplacians I and II, *Annales Institut Fourier* (1983).

[13] Donnelly H., On the Point Spectrum for Finite Volume Symmetric Spaces of Negative Curvature, *Comm. in P.D.E.* **6** (1981), pp. 963–992.

[14] Donnelly H., On the Cuspidal Spectrum for Finite Volume Symmetric Spaces, *J. Diff. Geom.* **17** (1982), pp. 239–253.
[15] Hirzebruch F., Hilbert Modular Surfaces, *L'Enseignment Mathématique* **19** (1973), pp. 183–281.
[16] Lax P., and Phillips R., Scattering Theory for Automorphic Functions, *Annals of Math. Studies* **87** (1976), Princeton Univ. Press.
[17] Müller W., Analytic Torsion and *R*-Torsion of Riemannian Manifolds, *Advances in Math.* **28** (1978), pp. 233–305.
[18] Müller W., *Spectral Theory of Non-Compact Riemannian Manifolds with Cusps and a Related Trace Formula*, preprint IHES, 1981.
[19] Müller W., Spectral Theory for Riemannian Manifolds with Cusps and a Related Trace Formula, *Math. Nachrichten* **111** (1983), pp. 197–288.
[20] Müller W., *Spectral Theory for Riemannian Manifolds with Cusps*, preprint, 1983.
[21] Müller W., *Signature Defects of Cusps of Hilbert Modular Varieties and Values of L-Series at* $s = 1$, preprint, 1983.
[22] Regge T., General Relativity without Coordinates, *Il Nuovo Cimento* **19** (1961), pp. 551–571.
[23] Selberg A., Harmonic Analysis and Discontinuous Groups in Weakly Symmetric Riemannian Spaces with Applications to Dirichlet Series, *J. Indian Math. Soc.* **20** (1956), pp. 47–87.
[24] Selberg A., On the Estimation of Fourier Coefficients of Modular Forms, *Proc. Symp. Pure Math. VIII*, A.M.S., Providence, 1965, pp. 1–15.
[25] Venkov A. B., The Spectral Theory of Automorphic Functions, the Selberg Zeta Function and Some Problems of Analytic Number Theory and Math. Physics, *Uspekhi Math. Nauk* **34** (1979), pp. 69–135.
[26] Zucker S., Hodge Theory with Degenerating Coefficients: L^2-Cohomology in the Poincaré Metric, *Annals of Math.* **109** (1979), pp. 415–476.
[27] Zucker S., L^2-Cohomology of Warped Products and Arithmetic Groups, *Inventiones Math.* (1983).

AKADEMIE DER WISSENSCHAFTEN DER DDR,
INSTITUT FÜR MATHEMATIK
DDR-1080 BERLIN,
MOHRENSTRASSE 39

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

RICHARD M. SCHOEN

Minimal Surfaces and Positive Scalar Curvature

An old idea in differential geometry going back more than 100 years is the idea of the interplay between curvature and geodesics via the second variation of arclength. This idea was used by Bonnet to bound the diameter of a surface with positive curvature, and this bound was extended to higher dimensions by B. Myers under the assumption of positive Ricci curvature. It was used by J. L. Synge to discuss the fundamental group of manifolds of positive sectional curvature. Moreover, much of the modern theory of Riemannian geometry is based on delicate refinements of this idea. Up until the present it has not been possible to use goedesic methods to get information concerning the geometry and topology of manifolds with positive scalar curvature. Strong restrictions on such manifolds were found by Lichnerowicz [7] who combined the index theorem of Atiyah–Singer with a vanishing theorem for solutions of the Dirac equation to show that such metrics do not exist on compact spin manifolds with nonzero $\hat{A}$-genus. N. Hitchin [5] then observed that the vanishing theorem of Lichnerowicz implies vanishing of other K-O characteristic numbers.

In the late sixties the positive mass problem in general relativity was posed to differential geometers. A special case of this problem involved understanding properties of three-dimensional manifolds of positive scalar curvature. The question of whether the n-dimensional torus admits a metric of positive scalar curvature was posed by Kazdan and Warner [6] in 1975. At that time it was not known how to apply the Lichnerowicz argument to manifolds such as T^n. Finally in 1977, S. T. Yau and the author [8, 11] solved the positive mass conjecture. The proof was based on the use of minimal hypersurfaces and relied heavily on the second variation of area. Since that time we have refined these arguments (see [10, 12, 13]) to gain a considerable understanding of manifolds of positive scalar curvature.

Our first results [13] concerned three-dimensional manifolds where we showed that if M is compact and $\Pi_1(M)$ contains a subgroup isomorphic to $\Pi_1(\Sigma)$ for a closed two-dimensional surface Σ of genus at least one, then M cannot carry a metric of positive scalar curvature. This gave a rather good understanding of the three-dimensional case; in fact, it shows that a compact three-manifold of positive scalar curvature must be a connected sum of copies of $S^2 \times S^1$, manifolds with finite fundamental group, and $K(\Pi, 1)$ manifolds where Π contains no surface group. It is a conjecture in topology that this last class of manifolds is empty. A short time later we were able to show directly that the $K(\Pi, 1)$ factors do not occur, and hence every compact three-manifold with positive scalar curvature is a connected sum of copies of $S^2 \times S^1$ and copies of compact manifolds with finite fundamental group. (The removal of the topological assumption mentioned above was also observed by Gromov and Lawson [4] using the spinor method.) It is also a topological conjecture that a compact three-manifold with finite fundamental group is the quotient of S^3 by a finite subgroup of SO(4), and hence admits a metric of constant positive sectional curvature. If this is true, then the structure of compact three manifolds with positive scalar curvature will be completely understood because a simple construction (see [10], [2]) shows that the connected sum of manifolds of positive scalar curvature also admits such a metric. In [13] we used stronger rigidity results for minimal surfaces obtained by D. Fischer-Colbrie and the author [1] to extend our program to complete noncompact three manifolds. In particular, we have shown that they are connected sums of $S^2 \times S^1$, manifolds of finite fundamental group, $\boldsymbol{R}^3$, and handlebodies. It can be seen that handlebodies admit complete metrics of positive (but not uniformly positive) scalar curvature. In particular, the only remaining question in three dimensions is whether a compact M^3 with $\Pi_1(M)$ finite necessarily admits a metric with positive scalar curvature.

In the autumn of 1978 we extended our methods to higher dimensions where we were able to establish [12] the positive action conjecture of S. Hawking. Again, as a byproduct of this investigation we were able to obtain substantial information about higher-dimensional manifolds of positive scalar curvature. In [10] we showed that a large class of manifolds, including T^n, cannot carry metrics of positive scalar curvature. The condition required in [10] was that the manifold have enough intersecting codimension one homology classes. The main new idea of [10] which enabled us to get results in higher dimensions was the idea of using the first eigenfunction of the second variation operator on a stable minimal hypersurface $\Sigma^{n-1} \subset M^n$ to conformally deform the induced metric of Σ to a metric of positive scalar curvature. We could then work inductively to prove the results. We did require in [10] the restriction $n \leqslant 7$ due to the possibility of singularities in high-dimensional area minimizing hypersurfaces. In the meantime we have been able to remove this technical condition and hence to apply our program in all dimensions. A short time after our results of [10], Gromov and Lawson [3] were able to extend the Lichnerowicz method to get results in all dimensions for a class of manifolds closely related to those considered in [10]. Over the past three years we have improved and localized the method of [10] (see [13] for the three-dimensional case) to prove a local theorem for positive scalar curvature which is analogous to the Bonnet–Myers' theorem. The result says that a manifold M^n of uniformly positive scalar curvature can be large in at most $n-2$ directions. A special case of this shows that compact enlargeable manifolds (note that this is the notion of enlargeability defined in [4] and does not require the fundamental group to be residually finite as considered in [3]) do not admit metrics of positive scalar curvature. The method can also be modified to show that a compact M^n which represents a homology class in a complete manifold N of nonpositive curvature cannot carry a metric of positive scalar curvature. This result was proven under the assumption that M is spin and N is compact by Gromov and Lawson [4]. An extension of this result shows that given a map $f\colon M^n \to N$ and a real cohomology class $\alpha \in H^k(N, \boldsymbol{R})$ such that $f^*(\alpha)$ is nonzero in $H^k(M, \boldsymbol{R})$, there exists a submanifold $M_1^{n-k} \subset M$ which intersects the dual homology class of $f^*(\alpha)$ nontrivially such that M_1 has a metric of positive scalar curvature and has trivial normal bundle in M.

Added in proof: Recently the author and S. T. Yau have extended their method to show that a compact four-dimensional $K(\Pi, 1)$ manifold cannot carry a metric of positive scalar curvature. We expect that more technical refinements of the same arguments will show this theorem to hold in arbitrary dimensions.

References

[1] Fischer-Colbrie D. and Schoen R., The Structure of Complete Stable Minimal Surfaces in 3-Manifolds of Nonnegative Scalar Curvature, *Comm. Pure and Appl. Math.* **33** (1980), pp. 199–211.

[2] Gromov M. and Lawson H. B., The Classification of Simply-Connected Manifolds of Positive Scalar Curvature, *Ann. of Math.* **111** (1980), pp. 423–434.

[3] Gromov M. and Lawson H. B., Spin and Scalar Curvature in the Presence of a Fundamental Group, *Ann. of Math.* **111** (1980), pp. 209–230.
[4] Gromov M. and Lawson H. B., *Positive Scalar Curvature and the Dirac Operator on Complete Riemannian Manifolds*, preprint.
[5] Hitchin N., Harmonic Spinors, *Adv. in Math.* **14** (1974), pp. 1–55.
[6] Kazdan J. and Warner F., Prescribing Curvatures, *Proc. Symp. in Pure Math.* **27** (1975), pp. 309–319.
[7] Lichnerowicz A., Spineurs harmoniques, *C. R. Acad. Sci. Paris, Ser. A-B* **257** (1963), pp. 7–9.
[8] Schoen R. and Yau S. T., Incompressible Minimal Surfaces, Three-Dimensional Manifolds with Nonnegative Scalar Curvature, and the Positive Mass Conjecture in General Relativity, *Proc. Nat. Acad. Sci.* **75** (6), (1978), p. 2567.
[9] Schoen R. and Yau S. T., Existence of Incompressible Minimal Surfaces and the Topology of Manifolds with Nonnegative Scalar Curvature, *Ann. of Math.* **110** (1979), pp. 127–142.
[10] Schoen R. and Yau S. T., On the Structure of Manifolds with Positive Scalar Curvature, *Manuscripta Math.* **28** (1979), pp. 159–183.
[11] Schoen R. and Yau S. T., On the Proof of the Positive Mass Conjecture in General Relativity, *Comm. Math. Phys.* **65** (1979), pp. 45–76.
[12] Schoen R. and Yau S. T., Proof of the Positive Action Conjecture in Quantum Relativity, *Phys. Rev. Let.* **42** (9) (1979), pp. 547–548.
[13] Schoen R. and Yau S. T., Complete Three-Dimensional Manifolds with Positive Ricci Curvature and Scalar Curvature, *Ann. of Math. Studies* **102** (1982), pp. 209–228.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

LEON SIMON

Recent Developments in the Theory of Minimal Surfaces

In this article we describe recent (within the past 4 or 5 years) results in minimal surface theory, with particular emphasis on the existence and regularity theory. For a more general survey, including some technical background discussion and a more comprehensive bibliography, the reader is referred to [30].

Throughout the article $\mathscr{H}^k$ will denote k-dimensional Hausdorff measure (in Euclidean space or in a general Riemannian manifold) and $B_\varrho^{(n)}(\xi)$ (often denoted simply $B_\varrho(\xi)$) will denote the ball in $\boldsymbol{R}^n$ with radius $\varrho > 0$ and centre ξ.

For convenience we break our discussion into several sections, although naturally there is some overlap between these.

1. 2-dimensional existence and regularity results

There have been a number of interesting recent developments in existence and regularity theory for 2-dimensional minimal surfaces in 3-dimensional manifolds.

First we want to describe the main regularity result of [5] concerning sequences $\{M_k\}$ of embedded discs in $\boldsymbol{R}^3$ (the extension from $\boldsymbol{R}^3$ to a general Riemannian 3-manifold is explicitly discussed in [18 § 4]), where $\{M_k\}$ is minimizing in the sense that there is $\{\varepsilon_k\} \downarrow 0$ with $\mathscr{H}^2(M_k) \leqslant \mathscr{H}^2(M) + \varepsilon_k$ whenever $k \geqslant 1$ and M is an embedded disc with $\partial M = \partial M_k$. In [5] it is proved that if $\xi \in \boldsymbol{R}^3$, if

$$\liminf_{k\to\infty} \mathscr{H}^2\big(M_k \cap B_\delta(\xi)\big) > 0 \qquad \forall \delta > 0,$$

$$\limsup_{k\to\infty} \mathscr{H}^2\big(M_k \cap B_\sigma(\xi)\big) < \infty \quad \text{ for some } \sigma > 0,$$

and if

$$\liminf_{k\to\infty} \operatorname{dist}(\xi, \partial M_k) > 0,$$

then, near ξ, some subsequence $\{M_{k'}\}$ of $\{M_k\}$ converges to an integral multiple of a smooth embedded minimal surface Σ in the sense that

$$\lim \int_{M_{k'}} f = n \int_{\Sigma} f \quad \forall f \in C_0(B_\tau(\xi))$$

for some integer $n \geqslant 1$ and $\tau > 0$.

This theorem has a number of very interesting consequences — it was originally proved (in [5]) to discuss existence of embedded solutions of the fixed genus Plateau problem (see [17] for another approach to the genus 0 case). More recently there have been other applications. For example the result is central in the work [18], where it is proved that one can find least-area representatives for the isotopy class of a given 2-dimensional surface (of any given genus) in a compact Riemannian 3-manifold. Another application is in the recent proof [29] (see also [27]) that there always exists an embedded minimal 2-sphere in the 3-sphere equipped with an arbitrary metric. Actually the work of [29] uses also, in a very essential way, modifications of the mini-max existence techniques of J. Pitts [19] (see § 2 below). The work of [29] has been extended to a higher genus setting by Pitto and Rubinstein [21]. Independently of [29], B. White [36] was able to establish that there are at least 4 distinct minimal 2-spheres in case the metric of S^3 is sufficiently close to the standard metric.

Concerning other 2-dimensional existence work, we want to mention particularly the work [13], in which it is proved that if Σ is an incompressible surface with least area in its homotopy class (see [26], [22], [17] concerning the existence of such least area representatives) and if the homotopy class of Σ contains an embedded surface, then Σ itself is embedded.

Concerning other regularity results, we mention the work [24] on curvature estimates and Bernstein Theorems for simply connected embedded minimal surfaces in 3-manifolds (assuming an *a priori* area bound), and the work [11], [23] concerning similar results in the stable immersed case (assuming no *a priori* area bound).

Finally we mention the very interesting work of Choi and Wang [9] on Poincaré inequalities and area bounds for embedded minimal surfaces in 3-manifolds. Using, in part, these results, Choi and Schoen [8] have proved the family of embedded minimal surfaces of a given genus is compact in the smooth topology. This leaves open whether or not such a family is finite.

2. Regularity of stable embedded hypersurfaces and the existence theory of J. Pitts

Suppose U is open in $\boldsymbol{R}^{n+1}$ and M is a stable embedded minimal hypersurface in U with $(\operatorname{sing} M)\cap U$ ($\equiv$ (closure $(M \sim M)\cap U$)) having zero $(n-2)$-dimensional Hausdorff measure. Schoen and the author [25] were recently able to prove, for such hypersurfaces, a regularity theorem analogous to the basic regularity theorems of De Giorgi [10] and Allard [3], but without any assumption requiring that the hypersurface be locally close in the *appropriate measure theoretic sense* to *one copy* of the n-dimensional disc; in [25] it is merely required that the hypersurface be locally sufficiently close in the *distance sense* to an n-dimensional disc. Using this result in combination with standard arguments from Geometric Measure Theory, the authors in [25] were able to obtain curvature estimates for M in U in case $n \leqslant 6$ (extending previous work of Schoen, Simon and Yau [28]). For $n \geqslant 7$ it is also proved in [25] that $\operatorname{sing} M \cap U$ has Hausdorff dimension at most $n-7$, and a compactness theorem for sequences of such hypersurfaces is proved in all dimensions $n \geqslant 2$.

The above results are important in the recent very striking work of J. Pitts [19] on existence of codimension 1 minimal submanifolds of oriented Riemannian manifolds. Pitts' method was first to use a mini-max technique on non-trivial paths representable by discrete homotopies in the space of n-dimensional integral cycles (equipped with the flat metric topology), in combination with a very clever local deformation argument, in order to prove existence of stationary integral varifolds having a strong "almost minimizing" property in sufficiently small annular neighbourhoods of each point of their support. Using this almost minimizing property in combination with the curvature estimates of [28], Pitts was then able to prove regularity of his almost minimizing varifolds, thus proving the required existence result, in case $n \leqslant 5$. Using the new regularity and compactness results referred to above, Schoen and the author [25] were able to extend Pitts' work to arbitrary dimension n. (For $n = 7$, the corresponding minimal hypersurface may have only isolated singularities and for $n \geqslant 8$ there may be a codimension 7 singular set.)

3. Almgren's work on regularity of arbitrary codimension area minimizing currents

Recently in [2], F. Almgren announced the very important result that, away from the support of its boundary, an n-dimensional area minimizing integral current (of arbitrary codimension) has a singular set of

Hausdorff dimension at most $n-2$. Prior to this a regularity theory in higher codimension was available only for mod 2 minimizing currents (which, roughly speaking, do not necessarily carry an orientation and are minimizing relative to both oriented and non-oriented comparisons.) (See [12], [21].)

Almgren's work relies on some very interesting new ideas and constructions; in particular he has developed a notion of multi-valued Dirichlet minimizing functions and a strong regularity theory for them.

The complex varieties in C^n show that Almgren's work is best possible for its type. The actual *structure* of the singular set is still open — for instance it is not even known whether or not the singular set always has *integer* Hausdorff dimension.

4. The tangent cone problem

Concerning the tangent cone problem, there have been 2 recent developments. First, B. White [34] in his Princeton thesis proved uniqueness of tangent cones for 2-dimensional area minimizing integral currents. Secondly, the author [31], in general dimensions, proved uniqueness of tangent cones for a general stationary varifold when at least one of the tangent cones is multiplicity 1 (and regular) away from 0. Almgren and Allard [1] had previously established this under an additional restriction, (unfortunately difficult to check and not generally satisfied) concerning "integrability" of Jacobi fields of the intersection of the cone with the unit sphere.

The work of [31] relies on a new (and quite generally applicable) theorem concerning a class of second order non-linear evolution equations of elliptic or parabolic type. (The same general theorem can be used to discuss, for example, uniqueness of tangent maps for harmonic mappings.)

5. The Plateau problem

B. White [35] recently made the important observation that if $n \geqslant 3$, if Γ is an $(n-1)$-dimensional compact submanifold of $\boldsymbol{R}^{n+1}$ diffeomorphic to S^{n-1} and if M (not necessarily diffeomorphic to a ball) is a smooth area minimizing solution of the unrestricted topological type oriented Plateau problem with boundary Γ (see [15] for discussion of the existence of such M), then there is a sequence $\{M_k\}$ of immersions of the *ball* with $\partial M_k = \Gamma \quad \forall k$ and such that $\{M_k\}$ converges weakly (in the sense of currents) to M and $\lim \mathscr{H}^n(M_k) = \mathscr{H}^n(M)$. This seems to essentially dash any hopes of developing a "fixed topological type" theory for the Plateau problem in dimension greater than 2.

Concerning the *non-parametric* Plateau problem, we mention the recent work [32] in which it is shown that solutions may have discontinuities at the "exceptional" boundary points where the mean curvature of the domain vanishes even if the given data is smooth. (For previous regularity work concerning this problem see [33].)

6. Examples of minimal submanifolds

Apart from the new examples already referred to above, we here want to mention the work of Anderson [4] concerning complete minimal submanifolds in hyperbolic space having specified boundary in the sphere at infinity, the new equivariant examples of Hsiang [16] (including some which are embeddings of S^n into S^{n+1}), and the examples of (none-cone) minimal submanifolds with isolated singularities [7], including some codimension 1 minimizing examples [14].

References

[1] Almgren F., Allard W., On the radial behaviour of minimal surfaces and the uniqueness of their tangent cones, *Annals of Math.* **113** (1981), pp. 215–265.

[2] Almgren F. J., *Q-valued functions minimizing Dirichlet's integral and the regularity of area minimizing rectifiable currents up to codimension 2*, Preprint of Research Announcement, 1983.

[3] Allard W. K., First variation of a Varifold, *Annals of Math.* **95** (1972), pp. 417–491.

[4] Anderson M. T., Complete minimal varieties in hyperbolic space, *Inventiones Math.* **69** (1982), pp. 477–494.

[5] Almgren F., Simon L., Existence of embedded solutions of Plateau's problem, *Ann. Scuola Norm. Sup. Pisa* **6** (1979), pp. 447–495.

[6] Bombieri E., Guisti E., Harnack's inequality for elliptic differential equations on minimal surfaces, *Invent. Math.* **15** (1972), pp. 24–46.

[7] Caffarelli L., Hardt R., Simon L., *Minimal surfaces with isolated singularities*, Preprint 1983.

[8] Choi H. I., Schoen R., to appear.

[9] Choi H. I., Wang A. N., *A first eigenvalue estimate for minimal surfaces*, Preprint.

[10] De Giorgi E., Frontiere orientate di misura minima, *Sem. Mat. Scuola Norm. Sup. Pisa* (1961), pp. 1–56.

[11] Fischer-Colbrie D., Schoen R., *The structure of complete stable minimal surfaces in 3-manifolds of non-negative scalar curvature*, Preprint.

[12] Federer H., The singular sets of area minimizing currents, *Bull. AMS* **76** (1970), pp. 767–771.

[13] Freedman M., Hass J., Scott P., *Least area incompressible surfaces in 3-manifolds*, Preprint.
[14] Hardt R., Simon L., to appear.
[15] Hardt R., Simon L., Boundary regularity and embedded solutions for the oriented Plateau problem, *Ann. of Math.* **110** (1979), pp. 439–486.
[16] Hsiang Wu-Yi, Preprint.
[17] Meeks W., Yau S.-T., The classical Plateau problem and the topology of 3-manifolds, *Topology* (1982).
[18] Meeks W., Simon L., Yau S.-T., Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature, *Annals of Math.* **116** (1982), pp. 621–659.
[19] Pitts J., Existence and regularity of minimal surfaces in Riemannian manifolds, *Mathematical Notes P.U.P.* 1981.
[20] Reifenberg E., *Annals of Math.* **80** (1964), pp. 1–21.
[21] Rubinstein H., to appear.
[22] Sacks J., Uhlenbeck K., Minimal immersions of closed Riemannian surfaces, Preprint.
[23] Schoen R., Estimates for stable minimal surfaces in 3-dimensional manifolds, *Annals of Math. Studies* **103** (1983).
[24] Schoen R., Simon L., Regularity of simply connected surfaces with quasiconformal Gauss map. To appear in: *Annals of Math. Studies Vol. on Minimal Submanifolds* (1983).
[25] Schoen R., Simon L., Regularity of stable minimal hypersurfaces, *Comm. Pure Appl. Math.* **34** (1981), pp. 741–788.
[26] Schoen R., Yau S.-T., Existence of incompressible minimal surfaces and the topology of 3-manifolds with nonnegative scalar curvature, *Annals of Math.* **110** (1979), pp. 127–142.
[27] Smith F., Thesis, University of Melbourne 1982.
[28] Schoen R., Simon L., Yau S.-T., Curvature estimates for minimal hypersurfaces, *Acta Math.* **134** (1975), pp. 275–288.
[29] Simon L., Smith F., to appear.
[30] Simon L., Survey lectures on minimal submanifolds, *Annals of Math. Studies* **103** (1983).
[31] Simon L., *Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems*, Preprint, Australian National University, 1983.
[32] Simon L., Boundary behaviour of solutions of the non-parametric least area problem, *Bull. Australian Math. Soc.* **26** (1982), pp. 17–28.
[33] Simon L., Boundary regularity of solutions of the non-parametric least area problem, *Annals of Math.* **103** (1976), pp. 429–455.
[34] White B., Thesis, Princeton University (1981).
[35] White B., to appear in: *Annals of Mathematics* (1983).
[36] White B., Research Announcement, *Notices AMS* **3** (1982), 300.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

KAREN K. UHLENBECK

Variational Problems for Gauge Fields

This talk will be a brief survey of recent results in the study of elliptic systems of partial differential equations of gauge field theory. The equations of this talk are the classical Euler–Lagrange equations corresponding to current quantum field theory models. It is not clear to what extent any of these results is interesting in physics; however, these equations coming from physics are now very important in mathematics itself. At the Congress in 1978, Arthur Jaffe gave a talk on gauge theories which included some classical theory [11]. Progress in the last five years can be measured by reading his article. The main progress has been the step from considering special, scalar forms of the equations to handling the full gauge theory. Even more remarkable are the applications which are turning up in pure mathematics. Special note should be made of the contributions of Clifford Taubes and Simon Donaldson to the subject.

The Yang–Mills–Higgs equations are the Euler–Lagrange equations for an action integral

$$S(\Phi, A) = \frac{1}{2}\int_M (|F_A|^2 + |D_A\Phi|^2 + V|\Phi|^2)\,d\mu.$$

Here M is a Riemannian n-dimensional manifold, and the objects which are variable are the connection A in a fixed principal bundle P over M with compact structure group G and a section Φ of an associated vector bundle $\eta = P\times_G E$ (Φ is called the Higgs field). The curvature of A is F_A, D_A is the covariant derivative induced by A on η, $V\colon \boldsymbol{R}^+\to\boldsymbol{R}$ is a potential function (which is usually taken to be $V(\sigma) = \lambda(1-\sigma)^2$). The Euler–Lagrange equations are of the form

$$D_A^*F_A = J = [D_A\Phi, \Phi],$$

$$D_A^*D_A\Phi + V'(|\Phi|^2)\Phi = 0.$$

The important technical aspect of gauge theory is its invariance under the group of gauge transformations. This is the group of sections $\mathscr{G} = C^\infty(M, P \times_{\mathrm{Ad}G} G)$. This must be divided out in some way, in each subject — geometry, partial differential equations and in physics — before gauge theory problems can be understood. Section 1 describes how this has been done successfully for partial differential equations. Section 2 describes results on the pure Yang–Mills fields over compact Riemannian manifolds, and in Section 3 we outline the monopole problem ($M = \boldsymbol{R}^3$, $G = \mathrm{SU}(2)$ and $V = 0$).

§ 1. Gauge problems

The group of gauge transformations plays the role in gauge field theory analogous to the diffeomorphism group for Einstein's equations. Luckily the gauge group is much easier to handle technically. The optimal local gauge fixing theorem is proved by a standard argument on openness and closedness of solutions to a non-linear PDE. Here L_k^p denotes the Sobolev space with k derivatives in L^p. Also $B_\sigma(x) = \{y \in M\colon\ \delta(x, y) \leqslant \sigma\}$.

THEOREM 1 [30]. *If $B_\sigma^n \subseteq M \subseteq \boldsymbol{R}^n$ and $A \in L_1^{n/2}(B_\sigma^n)$ and $\int_{|x|<\sigma} |F_A|^{n/2}\, dx \leqslant \varepsilon(n, G)$, where ε is a constant depending on the underlying domain dimension n and the group G, then the connection A can be written in local coordinates as $d + a$, where*

(a) $\operatorname{div} a = 0$,

(b) $(a|\partial B_\sigma^n)$ *normal* $= 0$,

(c) $\int_{|x|<\sigma} |\operatorname{grad} a|^{n/2} dx \leqslant K(n, G) \int_{|x|<\sigma} |F_A|^{n/2} dx$.

In other words, the curvature alone can be used to locally control the connection form, provided it is small enough in the $L^{n/2}$ norm. There are a number of elegant *global* slice theorems as well as this local theorem [2], [16]. Such a theorem in itself is not important, but its consequences are basic first steps in understanding gauge fields. We find immediately that every solution to the Yang–Mills equations with $A \in L_1^{n/2}$ is gauge equivalent to an analytic solution. This holds for coupled equations under suitable restrictions on the Higgs field [12], [16], [23].

Another corollary of the local gauge fixing theorem is a global compactness theorem.

THEOREM 2 [30]. *If M is compact, then the set of connections in L_1^p with $\int_M |F_A|^p d\mu \leqslant B$ for $p > n/2$ is weakly compact.*

Interesting phenomena occur for $p = n/2$ [21] and when M is non-compact [27]. Further gauge-fixing theorems are special to the removeable singularities theorems. See for example Uhlenbeck [28], [29], Parker [16] and Sibner [22], [23]. Global methods of constructing a gauge satisfying $d^* a = 0$ without the curvature condition are given in work of R. Schoen and myself [20].

§ 2. The Yang–Mills equations

The most spectacular application of gauge theory in mathematics occurred last year when Simon Donaldson used gauge theory to prove a theorem in four manifold topology. Over a four manifold, the second order Yang–Mills equation $D_A^* F_A = 0$ is a consequence of a first order system of partial differential equations $^*F_A = F_A$, called the *self-dual Yang–Mills equations*. (This is defined only on a four manifold, since $^*F_A$ is usually an $(n-2)$-form and F_A a 2-form.) In a beautiful series of papers [25], [26], Clifford Taubes has shown in detail how to use approximate superposition and the implicit function theorem to construct solutions to this equation over any four manifold (twistor theory works over special four manifolds). Simon Donaldson was able to use Taubes' construction and the ideas behind the compactness theory to study the moduli space of solutions to the self-dual equations $\mathscr{M} = \{A: F_A = {}^*F_A\}/G$. The dimension of M is computed via the Atiyah–Singer index theorem [2].

THEOREM 3 (Donaldson) [4, 7]. *For M^4 smooth, compact, simply connected and $\int_{M^4} w \wedge w \geqslant 0$ for w a closed 2-form on M^4, we have*

(a) *For generic metrics $\mathscr{M} - \{p_1, \ldots, p_l\}$ is a smooth, oriented, non-empty five manifold where $l = \dim H^2(M, \mathbf{Z})$.*

(b) *In a neighborhood of p_a, M is a cone on CP^2.*

(c) *$\mathscr{M}$ has a collar $M^4 \times (0, \lambda)$ $\mathscr{M}$ and $\mathscr{M} - M^4 \times (0, \lambda)$ is compact.*

The striking result is the corollary. Using $\mathscr{M}$ as a cobordism, he shows that for such manifolds, the unimodular bilinear form (given by $\int w \wedge w$) which maps $H^2(M, \mathbf{Z}) \to \mathbf{Z}$ is diagonal. This result and Freedman's recent solution of the four-dimensional Poincaré conjecture [8] show there are many topological four manifolds which cannot have smooth structures. This is the basis of the discovery of the "fake $\mathbf{R}^4$" [7].

There are many more interesting results on Yang–Mills equations. The compactness theorem shows the existence of many solutions to the Yang–Mills equations over compact two and three manifolds. Atiyah and Bott have shown the usefulness of these equations over Riemann surfaces in studying the moduli space of stable vector bundles [1] and it seems likely this can be done over compact Kähler manifolds of any dimension [6], [13]. Discovering a use for the Yang–Mills equations over three-manifolds remains a challenge.

The pure Yang–Mills equations are remarkably similar to the equations for harmonic maps, with dimension two special for harmonic maps and dimension four special for Yang–Mills fields. Both have relationships with other conformally invariant problems and "borderline" phenomena which are currently very much studied [3], [15]. It seems very likely that there are many analogies between the theorems on harmonic maps in two dimensions and those on Yang–Mills fields in four dimensions. In particular, Sedlacek [21] shows that direct minimization techniques work to obtain minima for the Yang–Mills functional with the topological invariant $\alpha(P) \in H^2(M^4, \pi_1 G)$ prescribed, rather than complete data for P. This compares to the theorem on minimizing the energy on maps $s\colon M^2 \to N$, with $\pi_1(s)$ prescribed, which was proved by Lemaire [14], Schoen–Yau [20] and Sacks–Uhlenbeck [18]. We expect further partial existence results of this sort to become standard in the geometric calculus of variations.

§ 3. Monopoles

Many new and interesting ideas, problems and phenomena in partial differential equations occur in studying the Yang–Mills–Higgs functional on $\boldsymbol{R}^n$. For $n = 2$, $G = S^1$ and $E = C$ solutions of Yang–Mills–Higgs equations describe certain super conductors and for $n = 3$, $G = \mathrm{SU}(2)$ and $E = \mathrm{su}(2) = \boldsymbol{R}^3$, they give equations thought to describe monopoles. (Again, the relevance of the mathematics to actual experimental physics is unclear.) Taubes has done extensive work on these equations for very special V. The study of the equations for other choices of V remains completely open. Since we did not do justice to his work on self-dual Yang–Mills equations, we give a brief outline of some of his work on monopoles.

This can be described without bundles. Here

$$S(A, \Phi) = \frac{1}{2} \int_{\boldsymbol{R}^3} |F_A|^2 + |D_A \Phi|^2 + \lambda(1 - |\Phi|^2)^2 \, dx,$$

where

$$A = (A_1, A_2, A_3)\colon \boldsymbol{R}^3 \to \Lambda^1(\boldsymbol{R}^3) \times \mathrm{su}(2),$$

$$\Phi\colon \mathbf{R}^3 \to \mathrm{su}(2),$$

$$F_A = \left\{\frac{\partial}{\partial x^\alpha} A_\beta - \frac{\partial}{\partial x^\beta} A_\alpha + [A_\alpha, A_\beta]\colon \mathbf{R}^3 \to \Lambda^2(\mathbf{R}^3) \times \mathrm{su}(2)\right\},$$

$$D_A \Phi = \left\{\frac{\partial}{\partial x^\alpha} \Phi + [A_\alpha, \Phi]\colon \mathbf{R}^3 \to \mathrm{su}(2)\right\}.$$

Boundary conditions are given by

$$\lim_{|x|\to\infty} |\Phi(x)| = 1,$$

$$\text{degree}\left(\frac{\Phi}{|\Phi|}\colon \{x\colon |x| = \sigma\} = S^2_\sigma \to S^2\right) = k \quad \text{for } \sigma \geqslant \sigma_0.$$

Spherically symmetric solutions for $k = \pm 1$, $\lambda \geqslant 0$, have been known to physicists for some time. The zeros of the solution Higgs field Φ of the Yang–Mills–Higgs equations locate the position of monopoles (if degree $+$) and anti-monopoles (if degree $-$) which have been superimposed according to a non-linear superposition principle into a single solution.

The parameter $\lambda = 0$ is special, since then the second order Yang–Mills–Higgs equations are consequences of a set of first order equations $^*F_A = \pm D_A\Phi$. By using an approximate superposition principle and an implicit function theorem, Taubes obtained the first existence theorem. The sign of k can be reversed by reversing orientation on $\mathbf{R}^3$.

THEOREM 4 [12]. *For $k > 0$ there exists a smooth $3k$ parameter family of solutions to $F_A = D_A\Phi$ with k zeros located exactly at $x_a \in \mathbf{R}^3$, $1 \leqslant a \leqslant k$, provided $|x_a - x_\beta| \geqslant d$ for $a \neq \beta$.*

Later Taubes was able to show that the zeros and relative gauge positions give a full parameter space for the solution space (modulo the gauge group) as long as the zeros (monopoles) are widely separated. Twistor theory also later gave more complete information a full parameter space of solutions, but with less geometric information on the widely spaced monopole solutions [10, 31].

One should note that solutions to $^*F_A = \pm D_A\Phi$ are minima of S for topological reasons. For $k = 0$, Taubes was able to show by a direct minimax argument the existence of saddle (non-minimal) critical points for S [27]. This work was extended by Dave Groisser to cover $\lambda > 0$ but small [9]. For $\lambda = 0$, it is conjectured by Taubes that a full Morse theory holds, which would imply the existence of many solutions to the Yang–Mills–Higgs equations. Many more open problems, expecially for $\lambda \neq 0$, are given by Jaffe and Taubes [12].

References

[1] Atiyah M. and Bott R., Yang–Mills Equations over Riemann Surfaces, to appear in *Phil. Trans. Roy. Soc. London.*

[2] Atiyah M., Hitchin N., and Singer I., Self-Duality in 4-Dimensional Riemannian Geometry, *Proc. Roy. Soc. London A* **362** (1978), p. 425.

[3] Brezis H. and Nirenberg L., Positive Solutions of Non-Linear Elliptic Equations Involving Critical Sobolev Exponents, *Commun. Pure and Appl. Math.*

[4] Donaldson S., An Application of Gauge Theory to the Topology of 4-Manifolds, to appear in *J. Diff. Geo.*

[5] Donaldson S., A New Proof of a Theorem of Narasimham and Seshadri, preprint, Oxford University, 1982.

[6] Donaldson S., The Yang–Mills Equations over Kähler Manifolds, Thesis, Oxford University, 1983.

[7] Freed, Freedman and Uhlenbeck, *Seminar on Gauge Theories and Four-Manifolds*, MSRI preprint series, Berkeley, 1983.

[8] Freedman M., The Topology of Four-Dimensional Manifolds, *J. Diff. Geo.* **17** (1983), pp. 357–455.

[9] D. Groisser (in preparation).

[10] Hitchin N., Monopoles and Geodesics, *Commun. Math. Phys.* **83** (1982), pp. 579–602.

[11] Jaffe A., Introduction to Gauge Theories, *Proceedings ICM*, Helsinki, 1978.

[12] Jaffe A. and Taubes C., *Vortices and Monopoles*, Progress in Physics #2, Birkhäuser, Boston, 1981.

[13] Kobayashi S., Curvature and Stability of Vector Bundles, *Proc. Japan Academy* **58** A (1982), pp. 158–162.

[14] Lemaire L., Applications harmoniques de surfaces riemanniennes, *J. Diff. Geo.* **13** (1978), pp. 51–78.

[15] Lions P. L., On the Existence of Positive Solutions of Semilinear Elliptic Equations, *SIAM Review* **24** (1982), pp. 441–467.

[16] Parker T., Gauge Theories on 4-Dimensional Riemannian Manifolds, *Commun. Math. Phys.* **85** (1982), pp. 563–602.

[17] Price P., *A Monotonicity Formula for Yang–Mills Fields*, preprint.

[18] Sacks J. and Uhlenbeck K., The Existence of Minimal Immersions of 2-Spheres, *Ann. of Math.* **113** (1981), pp. 1–24.

[19] Schoen R. and Uhlenbeck K., A Regularity Theory for Harmonic Maps, *J. Diff. Geo.* **17** (1982), pp. 307–335.

[20] Schoen R. and Yau S. T., The Existence of Incompressible Minimal Surfaces and the Topology of Three-Dimensional Manifolds with Non-Negative Scalar Curvature, *Ann. of Math.* **110** (1979), 127–142.

[21] Sedlacek S., A Direct Method for Minimizing the Yang–Mills Functional over 4-Manifolds, *Commun. Math. Phys.* **86** (1982), pp. 515–528.

[22] Sibner L., Removeable Singularities of Yang–Mills Fields in $\mathbf{R}^3$, to appear in *Compositio Mathematica.*

[23] Sibner L. and Sibner R., Removeable Singularities of Coupled Yang–Mills in $\mathbf{R}^3$, to appear in *Commun. Math. Phys.*

[24] Taubes C., The Existence of Multi-Monopole Solutions for the Non-Abelian,

Yang–Mills–Higgs Equations for Arbitrary Simple Gauge Groups, *Commun. Math. Phys.* **80** (1981), pp. 343–367.
[25] Taubes C., Self-Dual Yang–Mills Connections on Non-Self Dual Four Manifolds, *J. Diff. Geo.* **17** (1982), pp. 139–170.
[26] Taubes S., Self-Dual Connections on 4-Manifolds with Indefinite Intersection Matrix (submitted to *J. Diff. Geo.*).
[27] Taubes C., The existence of a Non-Minimal Solution to the SU(2) Yang–Mills–Higgs Equations on $\boldsymbol{R}^3$, *Commun. Math. Phys.* **86** (1982), Part I: pp. 257–298, Part II: pp. 299–320.
[28] Uhlenbeck K., Removeable Singularities in Yang–Mills Fields, *Commun. Math. Phys.* **83** (1982), pp. 11–29.
[29] Uhlenbeck K., Chern Classes of Sobolev Connections, to appear in *Commun. Math. Phys.*
[30] Uhlenbeck K., Connections with L^p Bounds on Curvature, *Commun. Math. Phys.* **83** (1982), pp. 31–42.
[31] Ward R., Deformations of the Embedding of the SU(2) Monopole Solution in SU(3), *Commun. Math. Phys.* **86** (1982), pp. 437–448.

THE UNIVERSITY OF ILLINOIS IN CHICAGO, 61680
HARVARD UNIVERSITY, CAMBRIDGE, 02138

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

È. B. VINBERG

Discrete Reflection Groups in Lobachevsky Spaces

1

The construction of discrete groups of motions of n-dimensional Lobachevsky space L^n is reduced in principle to constructing their fundamental polyhedra, together with the generating motions superposing the $(n-1)$-dimensional faces pairwise. By means of this method, H. Poincaré [25] described all finitely generated discrete groups of motions of Lobachevsky plane. In the memoir on Kleinian groups [26] he outlined an analogous program for 3-dimensional Lobachevsky space.

Poincaré's method was strictly proved and generalized in different directions by A. D. Aleksandrov [1], B. Maskit [16], and others. Recently G. D. Mostow [19] used it brilliantly to construct several discrete groups of motions of the complex hyperbolic plane. However, in Lobachevsky spaces of dimensions $\geqslant 3$ this method was used successfully only for groups generated by reflections with respect to hyperplanes (so-called *reflection groups*) and to its subgroups of finite index. In other cases metric conditions on the fundamental polyhedron are usually so complicated and combinatorial conditions are so non-unique that constructing the required polyhedra seems almost hopeless.

It is more realistic, by using Poincaré's method, to check the discreteness of groups discovered for other reasons, for example, within the framework of Thurston's ideology [31].

2

Let us turn to reflection groups.

A convex polyhedron in Euclidean or Lobachevsky space is called a *Coxeter polyhedron* if all its dihedral angles are submultiples of π.

The construction of discrete reflection groups by Poincaré's method is reduced to constructing Coxeter polyhedra. Namely, the fundamental polyhedron of any discrete reflection group is a Coxeter polyhedron and, vice versa, any Coxeter polyhedron P is the fundamental polyhedron of some discrete reflection group Γ. The group Γ is generated by the reflections R_i with respect to the hyperplanes H_i bounding P. The defining relations of Γ are

$$R_i^2 = 1, \quad (R_i R_j)^{n_{ij}} = 1,$$

where $n_{ij} = \pi/(\text{the angle between } H_i \text{ and } H_j)$ if $P \cap H_i \cap H_j \neq \emptyset$ and $n_{ij} = \infty$ otherwise. Moreover, if $P \cap H_i \cap H_j = \emptyset$, then $H_i \cap H_j = \emptyset$.

It is convenient to describe a Coxeter polyhedron P by its *Coxeter diagram*. To each hyperplane H_i bounding P there corresponds a vertex v_i of the diagram. Two vertices v_i, v_j are joined as follows:

if H_i and H_j	then v_i and v_j are joined
intersect at an angle of π/n_{ij}	by an $(n_{ij}-2)$-tuple line or by a simple line with index n_{ij}
are parallel	by a thick line
diverge	by a dotted line

The Coxeter polyhedra belongs to the polyhedra whose dihedral angles do not exceed $\pi/2$. We shall call the latter ones *acute*.

Coxeter polyhedra in Euclidean spaces were classified in 1934 by H. S. M. Coxeter [6]. It was decisive for this classification that in the Euclidean space any acute polyhedron is the direct product of some number of simplices, a simplicial cone and a full space of some dimension.

The only known simple combinatorial property of acute polyhedra in Lobachevsky spaces is that they are *simple*. (A convex polyhedron is called *simple* if each k-codimensional face of it is contained in exactly k faces of codimension 1.)

It is unlikely that one could classify all Coxeter polyhedra in Lobachevsky spaces. However, the most interesting among them are the ones of finite volume, i.e., those which are the convex hulls of finite sets of points, ordinary or infinite. Classifying such Coxeter polyhedra seems to be an extremely difficult but solvable problem.

We shall expose what is known about this problem.

3

Coxeter polygons of finite area in Lobachevsky plane were described in 1882 by H. Poincaré [25] and W. Dyck [8]. Such a polygon may have an arbitrary number of sides and arbitrary angles $\pi/n_1, \ldots, \pi/n_k$, provided $1/n_1 + \ldots + 1/n_k < k-2$. Moreover, it depends on $k-3$ continuous parameters.

In Lobachevsky space of dimension $\geqslant 3$, a Coxeter polyhedron of finite volume is uniquely determined by its dihedral angles.

It follows, for instance, from the strong rigidity theorem for discrete subgroups of the group $O_{n,1}$ ([18], [27]). Consequently, there is at most countably many Coxeter polyhedra of finite volume in n-dimensional Lobachevsky space L^n for $n \geqslant 3$.

Examples of such polyhedra appeared in various works starting from the end of the last century.

In 1883 V. Schlegel [28] found all partitions of Lobachevsky spaces into equal bounded regular polyhedra. The symmetry group of such a partition is generated by reflections. Its fundamental polyhedron is a simplex. In 1950 F. Lannér [10] found all bounded Coxeter simplices in L^n. Not all of them result of the partitions into regular polyhedra, but they exist, as well as these partitions, only for $n \leqslant 4$. It is not so hard to enumerate also unbounded Coxeter simplices of finite volume in L^n. They exist only for $n \leqslant 9$ ([32], [5]).

In 1892 L. Bianchi [4], using a method due to R. Fricke, studied the discrete groups of motions of 3-dimensional Lobachevsky space which are represented by the unimodular (2×2)-matrices over the integers of the imaginary quadratic fields $Q(\sqrt{-d})$. He showed that, for $n \geqslant 19$, $n \neq 14, 17$, these groups are commensurable with some reflection groups, whose fundamental polyhedra he determined explicitly. These polyhedra are unbounded, but of finite volume. Some of them have a rather complicated combinatorial structure.

In 1931 F. Löbell [11] constructed some examples of 3-dimensional compact hyperbolic space forms. To do this, he considered a rectangular 14-hedron in L^3 (like dodecahedron but with hexagonal bases). His space forms were the factors of L^3 with respect to some subgroups of finite index of the reflection group associated with this 14-hedron.

In 1965–66 V. S. Makarov ([12]–[14]) proposed various geometrical methods for constructing Coxeter polyhedra in L^3. This enabled him to construct discrete reflection groups having a fundamental polyhedron of finite volume and containing elements of any prescribed order or prescribed plane reflection groups and so on. These works of Makarov and the interest in geometrical constructions of discrete groups, evoked by the problem of arithmeticity of discrete subgroups of semisimple Lie groups, stimulated research into hyperbolic discrete reflection groups.

4

In 1970 E. M. Andreev ([2], [3]) got an exhaustive description of acute polyhedra of finite volume in L^3 and, as a consequence, a *classification of Coxeter polyhedra of finite volume in* L^3. Namely, he showed some necessary and sufficient conditions under which in L^3 there exists a convex polyhedron of finite volume having an assigned combinatorial structure and assigned dihedral angles $\leqslant \pi/2$. For a polyhedron which is not a simplex these conditions have the form of simple linear inequalities. For instance, the sum of the dihedral angles between three faces having a vertex in common must be not less than π.

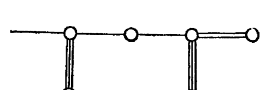
Such description of acute polyhedra is a peculiarity of 3-dimensional Lobachevsky space. For simplicity, let us consider only bounded polyhedra. It follows from the Euler equation that the number of degrees of freedom of a bounded simple 3-dimensional polyhedron of an assigned combinatorial structure equals the number of its dihedral angles (i.e., the number of its edges). For polyhedra of higher dimensions, the number of degrees of freedom may be less than the number of dihedral angles, and then the existence conditions cannot be reduced to inequalities. There are examples of bounded Coxeter polyhedra with such "superfluous rigidity" (see below). For this reason one cannot expect an Andreev-type classification of Coxeter polyhedra in the spaces L^n for $n \geqslant 4$.

5

Various examples of the Coxeter polyhedra of finite volume (both bounded and unbounded) in the spaces L^4 and L^5 were given in the works of the author [32], [33], V. S. Makarov [15] and I. M. Kaplinskaja [9].

In the space L^5 there is a curious "construction set" which enables us to construct infinitely many superrigid bounded Coxeter polyhedra [38].

The main piece of this set is cut out of the Coxeter polyhedron P of infinite volume with diagram



In the projective model of the space L^5, we have $P = \hat{P} \cap L^5$, where $\hat{P}$ is a convex polyhedron in the projective space which is combinatorially isomorphic to the direct product of the tetrahedron and the triangle. All the faces of $\hat{P}$ excluding three vertices intersect L^5. These three vertices do not belong even to the closure of L^5. Truncating them by their polar hyperplanes we obtain a bounded Coxeter polyhedron P' in the space L^5, which has three simplicial 4-dimensional faces of types

o≡o—o—o—o, o≡o—o—o=o and o≡o—o—o≡o,

orthogonal to the adjacent faces.

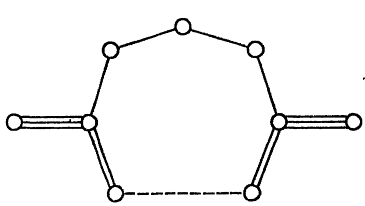
Putting together several copies of P' along equal simplicial faces one can construct infinitely many tree-like bounded Coxeter polyhedra. Moreover, in this construction set one may use two simplicial prisms

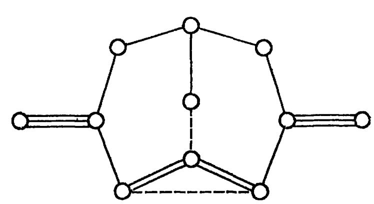
o≡o—o—o—o—o------o,

o≡o—o—o—o≡o------o,

discovered by Makarov [15], each of them has a base of type o≡o—o—o—o orthogonal to the lateral faces.

In the spaces L^6 and L^7 examples of bounded Coxeter polyhedra were constructed in an arithmetical way by V. O. Bugaenko. Their diagrams are as follows:



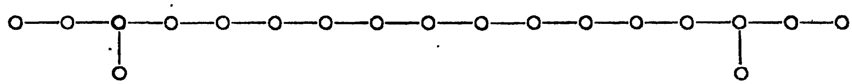


It is unknown whether there exist bounded Coxeter polyhedra in L^n for $n \geqslant 8$. However, the following theorem has been proved [36–38].

THEOREM 1. *In the space L^n for $n \geqslant 30$ there are no bounded Coxeter polyhedra.*

The proof is based on a recent result on combinatorics of simplicial convex polyhedra ([30], [7]) that is rephrased by duality for simple polyhedra. As V. V. Nikulin noted [22], it follows from this result that the mean complexity of a face of a fixed dimension of a bounded simple convex polyhedron tends to complexity of the cube as the dimension of the polyhedron becomes infinite. For the proof of Theorem 1, the above property is applied to 2-dimensional faces. In this case this means, roughly speaking, that any bounded simple convex polyhedron of high dimension has many tetragonal or trigonal 2-dimensional faces.

This method is not suitable for unbounded Coxeter polyhedra of finite volume, because the completion of such a polyhedron may be non-simple. Examples of such polyhedra in the space L^n are known for $n \leqslant 19$ ([33], [35], [39], [40]). The simplest of them for large n's is the polyhedron in L^{17} with the diagram



Its completion is combinatorially isomorphic to the pyramid over the direct product of two 8-dimensional simplices ([33], [35]).

It is unknown whether there exist Coxeter polyhedra of finite volume in Lobachevsky spaces of arbitrarily high dimension.

6

Many hyperbolic discrete reflection groups are obtained *in an arithmetical way*. Each arithmetical construction of such a group is tied to a totally real algebraic number field K [32]. We will consider only the case $K = \mathbf{Q}$.

Let L be a quadratic lattice of signature $(n, 1)$. Consider the projective model of n-dimensional Lobachevsky space associated with the pseudo-Euclidean space $E^{n,1} = L \otimes_{\mathbf{Z}} \mathbf{R}$. The group of its motions is a subgroup $O'_{n,1}$ of index 2 of the group $O_{n,1}$ of pseudo-orthogonal transformations of the space $E^{n,1}$. Reflections with respect to hyperplanes of Lobachevsky space are represented in this model as linear transformations of the form

$$R_e \colon x \mapsto x - \frac{2(e, x)}{(e, e)} e,$$

where $e \in E^{n,1}$, $(e, e) > 0$.

Let $O(L)$ denote the group of pseudo-orthogonal transformations of the space $E^{n,1}$ preserving L. Then $O'(L) = O(L) \cap O'_{n,1}$ is a discrete group of motions of Lobachevsky space. Its fundamental polyhedron has a finite volume [29]; it is bounded if and only if the lattice L is anisotropic (which is possible only for $n \leqslant 3$).

Denote by $O_r(L)$ the group generated by all reflections belonging to $O'(L)$. There is an effective algorithm to find the fundamental polyhedron $P(L)$ of this group ([33], [35]). The lattice L is called *reflective* if $[O(L):O_r(L)]<\infty$ or, equivalently, if the polyhedron $P(L)$ has a finite volume.

In the works [33]–[35], [40] (see also [17]) it was proved that *an unimodular quadratic lattice of signature* $(n,1)$ *is reflective if and only if* $n\leqslant 19$.

It is unknown whether there exists any reflective quadratic lattice of signature $(n,1)$ *for* $n\geqslant 20$. However, the following theorem has been proved ([36], [37]).

THEOREM 2. *There are no reflective quadratic lattices of signature* $(n,1)$ *for* $n\geqslant 30$.

Quadratic lattices L_1 and L_2 are called *similar* if there exist a group isomorphism $f\colon L_1\to L_2$ and a number $c\in Q$ such that $\big(f(x),f(y)\big)=c(x,y)$ for all $x,y\in L_1$.

THEOREM 3. (V. V. Nikulin [21], [22].) *For any* $n\geqslant 2$, *there are, up to similarity, only finitely many reflective quadratic lattices of signature* $(n,1)$.

Due to Theorems 2 and 3 we get the hope of classifying reflective quadratic lattices. However, it would be very laborious work.

7

Under the assumptions as above, denote by $O_r^{(2)}(L)$ the group generated by all reflections $R_e\in O'(L)$ for which $e\in L$, $(e,e)=2$. The lattice L is called *2-reflective* if $[O(L):O_r^{(2)}(L)]<\infty$.

To each *algebraic surface* X *of type* $K3$ there corresponds the quadratic lattice L_X of its algebraic cycles, the scalar product on it being defined as the intersection number with a minus sign. It is an even lattice of signature $(n,1)$, where $0\leqslant n\leqslant 19$. The automorphism group of X up to a finite central extension is described in terms of L_X [24]. In particular, it is finite if and only if the lattice L_X is 2-reflective.

V. V. Nikulin ([20], [23]) enumerated all 2-reflective even quadratic lattices of signature $(n,1)$ for $n\geqslant 4$. Later, the author of this note extended this classification to the case $n=3$ and Nikulin himself — to the case $n=2$. All the quadratic lattices found turned out to be the lattices of algebraic cycles of $K3$-surfaces. The maximal value of n for them was equal to 18.

If the lattice L_X is reflective but not necessarily 2-reflective, the theory of discrete reflection groups enables us to find the automorphism group of X and may also be applied to discovering projective models of X [39].

References

[1] Александров А. Д., О заполнении пространства многогранниками, *Вестник ЛГУ* **2** (1954), pp. 34–43.

[2] Андреев Е. М., О выпуклых многогранниках в пространствах Лобачевского, *Матем. сб.* **81** (1970), pp. 445–478.

[3] Андреев Е. М., О выпуклых многогранниках конечного объема в пространствах Лобачевского, *Матем. сб.* **83** (1970), pp. 256–260.

[4] Bianchi L., Sui gruppi de sostituzioni lineari con coefficienti appartenenti a corpi quadratici immaginari, *Math. Ann.* **40** (1892), pp. 332–412.

[5] Bourbaki N., *Groupes et algèbres de Lie*, ch. V, Hermann, Paris 1968.

[6] Coxeter H. S. M., Discrete groups generated by reflections, *Ann. Math.* **35** (1934), pp. 588–621.

[7] Данилов В. И., Геометрия торических многообразий, *Успехи матем. наук* **33** (1978), pp. 85–134.

[8] Dyck W., Gruppentheoretische Studien, *Math. Ann.* **20** (1882), pp. 1–45.

[9] Каплинская И. М., О дискретных группах, порожденных отражениями в гранях симплициальных призм в пространствах Лобачевского, *Матем. заметки* **15** (1974), pp. 159–164.

[10] Lannér F., On complexes with transitive groups of automorphisms, *Comm. Sém. Math. Univ. Lund* **11** (1950), pp. 1–71.

[11] Löbell F., Beispiele geschlossener dreidimensionaler Clifford–Kleinscher Räume negativer Krümmung, *Ber. Sächs. Akad. Wiss. Leipzig* **83** (1931), pp. 167–174.

[12] Макаров В. С., Об одном классе разбиений пространства Лобачевского, *Докл. АН СССР* **161** (1965), pp. 277–278.

[13] Макаров В. С., Об одном классе дискретных групп пространства Лобачевского, имеющих бесконечную фундаментальную область конечной меры, *Докл. АН СССР* **167** (1966), pp. 30–33.

[14] Макаров В. С., Об одном классе двумерных федоровских групп, *Изв. АН СССР, сер. матем.* **31** (1967), pp. 531–542.

[15] Макаров В. С., *О федоровских группах четырехмерного и пятимерного пространств Лобачевского, Исследования по общей алгебре*, Кишиневский ун-т, 1970, pp. 120–129.

[16] Maskit B., On Poincaré's theorem for fundamental polygons, *Adv. in Math.* **7** (1971), pp. 219–230.

[17] Meyer J., Präsentation der Einheitengruppe der quadratischen Form $F(X) = -X_0^2+X_1^2+\ldots+X_{18}^2$, *Arch. Math.* **29** (1977), pp. 261–266.

[18] Mostow G. D., *Strong rigidity of locally symmetric spaces*, Ann. of Math. Studies, **78**, Princeton Univ. Press 1973.

[19] Mostow G. D., On a remarkable class of polyhedra in complex hyperbolic space, *Pacific J. Math.* **86** (1980), pp. 171–276.

[20] Никулин В. В., О фактор-группах групп автоморфизмов гиперболических форм по подгруппам, порожденным 2-отражениями, *Докл. АН СССР* **248** (1979), pp. 1307–1309.

[21] Никулин В. В., Об арифметических группах, порожденных отражениями, в пространствах Лобачевского, *Изв. АН СССР, сер. матем.* **44** (1980), pp. 637–669.

[22] Никулин В. В., О классификации арифметических групп, порожденных отражениями, в пространствах Лобачевского, *Изв. АН СССР, сер. матем.* **45** (1981), pp. 113–142.

[23] Никулин В. В., О фактор-группах групп автоморфизмов гиперболических форм по подгруппам, порожденным 2-отражениями. Алгебро-геометрические исследования, *Современные проблемы математики,* т. 18, ВИНИТИ, 1981, pp. 3–114.

[24] Пятецкий-Шапиро И. И., Шафаревич И. Р., Теорема Торелли для алгебраических поверхностей типа $K3$, *Изв. АН СССР, сер. матем.* **35** (1971), pp. 530–572.

[25] Poincaré H., Théorie des groupes fuschsiennes, *Acta math.* **1** (1882), pp. 1–62.

[26] Poincaré H., Mémoire sur les groupes kleinéens, *Acta math.*, 3 (1883), pp. 49–92.

[27] Prasad G., Strong rigidity of Q-rank 1 lattices, *Invent. math.* **21** (1973), pp. 255–286.

[28] Schlegel V., Theorie der homogenen zusammengesetzten Raumgebilde, *Verh. K. Leopold.-Carolin Dtsch. Akad. Naturforsch.* **44** (1883), pp. 343–459.

[29] Siegel C., Einheiten quadratischer Formen, *Abh. Math. Sem. Hansischen Univ.* **13** (1940), pp. 209–239.

[30] Stanley R. P., The upper bound conjecture and Cohen–Macaulay rings, *Studies in Appl. Math.* **54** (1975), pp. 135–142.

[31] Thurston W. P., Three dimensional manifolds, Kleinian groups and hyperbolic geometry, *Bull. Amer. Math. Soc.* **6** (1982), pp. 357–382.

[32] Винберг Э. Б., Дискретные группы, порожденные отражениями, в пространствах Лобачевского, *Матем. сб.* **72** (1967), pp. 471–488.

[33] Винберг Э. Б., О группах единиц некоторых квадратичных форм, *Матем. сб.* **87** (1972), pp. 18–36.

[34] Винберг Э. Б., Об унимодулярных целочисленных квадратичных формах, *Функц. анализ* **6** (1972), pp. 25–32.

[35] Vinberg E. B., Some arithmetical discrete groups in Lobačevskiĭ spaces, *Proc. Int. Coll. on Discrete Subgroups of Lie Groups and Appl. to Moduli (Bombay, January 1973)*, Oxford University Press, 1975, pp. 323–348.

[36] Винберг Э. Б., Отсутствие кристаллографических групп отражений в пространствах Лобачевского большой размерности, *Функц. анализ* **15** (1981), pp. 79–80.

[37] Винберг Э. Б., Отсутствие кристаллографических групп отражений в пространствах Лобачевского большой размерности, *Труды Моск. матем. об-ва* **47** (1984), pp. 68–102.

[38] Винберг Э. Б., Гиперболические группы отражений, *Успехи матем. наук* (to appear).

[39] Vinberg É. B., The two most algebraic $K3$-surfaces, *Math. Ann.* **265** (1983), pp.1–21.

[40] Винберг Э. Б., Каплинская И. М., О группах $O_{18,1}(\mathbf{Z})$ и $O_{19,1}(\mathbf{Z})$, *Докл. АН СССР* **238** (1978), pp. 1273–1275.

**Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa**

О. Я. ВИРО

Успехи последних 5 лет в топологии вещественных алгебраических многообразий

Топология вещественных алгебраических многообразий как самостоятельная область исследований появилась немногим более 100 лет назад. Внимание математического сообщества было привлечено к ней Д. Гильбертом в 1900 году, когда он включил основные её задачи в свой знаменитый список проблем [7, 16-я проблема]. В течение последних 12 лет эта область развивалась особенно интенсивно вследствие привлечения новых топологических средств, начатого В. И. Арнольдом [1] и В. А. Рохлиным [19]. Несколько обзоров посвящены этому развитию. Его начальный этап (до 1974 года) отражен в обзоре Д. А. Гудкова [9]. Состояние топологии плоских кривых на 1978 год описано в обзорах Г. Вилсона [6] и В. А. Рохлина [21]. Топологии поверхностей был посвящен доклад В. М. Харламова [25] на предыдущем Международном Конгрессе Математиков.

Цель настоящего доклада — описать развитие предмета за последние 5 лет. Из-за ограничения объема здесь невозможно дать полный обзор. Я ограничусь следующими темами: (i) комплексные топологические характеристики неособых плоских проективных вещественных алгебраических кривых, см. §§ 1–5; (ii) комплексные топологические характеристики вещественных алгебраических поверхностей, § 6; (iii) классификация кривых и поверхностей относительно жестких изотопий, § 7; (iv) новые методы построения вещественных алгебраических многообразий с предписанными топологическими свойствами, § 8. Следующие темы были бы обязательно рассмотрены, если бы объем не был так ограничен: (i) индексы особенностей полиномиальных векторных полей, см. А. Г. Хованский [27]; (ii) новые ограничения на топологию неособых вещественных алгебраических поверхностей, найденные В. В. Никулиным [15] (в § 4 формулируются, однако, ограничения на топологию кривых, являющиеся следствиями этих ограничений); (iii) кривые на поверхностях и, в частности, на квадриках, см. Д. А. Гудков [10] и В. И. Звонилов [11]; (iv) особые кривые.

1. Вещественные алгебраические кривые как комплексные объекты

Простейшими и важнейшими алгебраическими многообразиями являются неособые плоские проективные вещественные кривые. Для краткости будем называть их просто *кривыми*. Множество вещественных точек кривой A будем обозначать через $\mathbf{R}A$. Оно является гладким замкнутым одномерным подмногообразием вещественной проективной плоскости $\mathbf{R}P^2$. Его компоненты гомеоморфны окружности. Если степень кривой четна, то все они располагаются в $\mathbf{R}P^2$ двусторонне, а если степень нечетна, то имеется ровно одна односторонняя компонента. Двусторонние компоненты называются *овалами*. Изотопический тип кривой $\mathbf{R}A \hookrightarrow \mathbf{R}P^2$ (или, что то же, топологический тип пары $(\mathbf{R}P^2, \mathbf{R}A)$) определяется схемой взаимного расположения овалов, которую называют также *вещественной схемой*.

В силу традиции, восходящей к Гильберту, главным вопросом топологии вещественных алгебраических кривых считался вопрос о том, какие вещественные схемы реализуются кривыми данной степени. Однако, уже Ф. Клейн [13] ставил вопрос шире. Он интересовался также тем, как вещественная схема кривой связана с расположением ее вещественной части $\mathbf{R}A$ в множестве $\mathbf{C}A$ ее комплексных точек.

$\mathbf{C}A$ есть ориентированное гладкое связное замкнутое двумерное подмногообразие комплексной проективной плоскости $\mathbf{C}P^2$, инвариантное относительно инволюции conj: $\mathbf{C}P^2 \to \mathbf{C}P^2$: $(z_0 : z_1 : z_2) \mapsto$ $\mapsto (\bar{z}_0 : \bar{z}_1 : \bar{z}_2)$. Кривая $\mathbf{R}A$ есть множество неподвижных точек сужения этой инволюции. Она может разбивать или не разбивать $\mathbf{C}A$. В первом случае A называется кривой *типа* I или *разбивающей* кривой, во втором случае — кривой *типа* II или *неразбивающей* кривой. В первом случае $\mathbf{R}A$ разбивает $\mathbf{C}A$ на две связные половины, естественные ориентации которых определяют на $\mathbf{R}A$, как на их общем крае, две противоположные ориентации. Эти ориентации называются *комплексными*. Вещественная схема кривой, обогащенная указанием типа кривой и, в случае типа I, комплексными ориентациями, называется *комплексной схемой* кривой.

Деление кривых на типы ввел Ф. Клейн [13]. Комплексные ориентации были введены в топологию вещественных алгебраических кривых Рохлиным [20]. Как я недавно узнал, они появились в работе И. Г. Петровского [17] о лакунах дифференциальных уравнений в частных производных. Комплексные схемы появились 5 лет назад в работе Рохлина [21]. В новейшем развитии предмета они занимают центральное положение. Вообще, в последние годы распространилось более широкое понимание задач топологии вещественных алгебраических многообразий, при котором в качестве основного объекта выступают вещественные многообразия вместе с их расположением в комплексификации, а не вещественные многообразия сами по себе, см. Рохлин [21].

Для ответа на вопрос, какой может быть вещественная и комплексная схемы кривой данной степени, необходима работа в двух направлениях: во-первых, необходимо найти ограничения, налагаемые алгебраической природой кривой на ее схемы; во-вторых, необходимо найти методы построения кривых данной степени с заданной схемой.

Значительная часть известных ограничений следует из сравнительно небольшого числа чисто топологических свойств алгебраических кривых. Поэтому наряду с алгебраическими кривыми полезно рассмотреть объекты, топологически иммитирующие их. Ориентированное гладкое связное замкнутое двумерное подмногообразие M комплексной проективной плоскости CP^2 будем называть *гибкой кривой степени* m, если: (i) оно реализует m-кратную образующую группы $H_2(CP^2)$; (ii) его род равен $(m-1)(m-2)/2$; (iii) оно инвариантно относительно conj; (iv) поле его касательных плоскостей на $M\cap \boldsymbol{R}P^2$ можно продеформировать в классе плоскостей, инвариантных относительно conj, в поле прямых пространства CP^2, касательных к $M\cap \boldsymbol{R}P^2$; (v) группа $\pi_1(CP^2\setminus M)$ коммутативна. (Возможно, это определение не является окончательным, будущие исследования могут привести, например, к добавлению новых условий.) С $\boldsymbol{R}P^2$ гибкая кривая пересекается по гладкому одномерному подмногообразию, которое называется ее вещественной частью. Очевидно, множество комплексных точек алгебраической кривой степени m является гибкой кривой степени m. Все сказанное выше об алгебраических кривых и их схемах переносится без изменений на случай гибких кривых. Будем говорить, что ограничения на схемы кривых степени m, которые доказаны для схем гибких кривых степени m, имеют *топологическое происхождение*.

2. Числовые характеристики и кодирование схем кривых

Чтобы формулировать ограничения, введём некоторые понятия и числовые характеристики, связанные со схемами. Говорят, что два овала образуют *инъективную пару*, если один из них охватывается другим. Совокупность h овалов, любые два из которых образуют инъективную пару, называется *гнездом* глубины h. Инъективная пара овалов кривой типа I называется *положительной*, если ориентации овалов, определяемые комплексной ориентацией, индуцируются ориентацией ограничиваемого ими кольца. Овалы кривой типа I нечетной степени делятся на положительные и отрицательные. Именно, рассмотрим ленту Мебиуса, остающуюся от $\mathbf{R}P^2$ после удаления внутренности овала. Если целочисленные гомологические классы, реализуемые в ней овалом и удвоенной односторонней компонентой с ориентациями, определяемыми комплексной ориентацией, различаются знаком, то овал называется *положительным*, в противном случае он называется *отрицательным*. В случае кривой типа I четной степени на положительные и отрицательные делятся только невнешние овалы. Именно, невнешний овал положителен, если он составляет с охватывающим его внешним овалом положительную пару, и отрицателен в противном случае. Овал называется *четным*, если он лежит внутри четного числа других овалов. Эйлерова характеристика компоненты дополнения кривой называется *характеристикой* овала, ограничивающего эту компоненту извне. Компонента дополнения кривой называется *четной*, если каждый из овалов, ограничивающих ее изнутри, охватывает нечетное число овалов.

Важнейшими числовыми характеристиками вещественной схемы являются следующие: l — число овалов, p — число четных овалов, n — число нечетных овалов, l^+, l^0 и l^- — числа овалов с положительными, нулевыми и отрицательными характеристиками; p^+, p^0, p^- и n^+, n^0, n^- — аналогичные числа четных и нечетных овалов; Π — число инъективных пар овалов, π и ν — числа четных и нечетных непустых овалов, ограничивающих извне четные компоненты дополнения кривой; h_r — наибольшее количество овалов, входящих в объединение $\leqslant r$ гнезд; h_r' — наибольшее количество овалов набора, содержащегося в объединении $\leqslant r$ гнезд и не содержащего овала, который охватывал бы все остальные овалы этого набора. Характеристиками *комплексной* схемы кривой *типа* I являются также: Π^+ и Π^- — числа положительных и отрицательных инъективных пар, и Λ^+ и Λ^- — числа положительных и отрицательных овалов.

Для описания вещественных схем применяется следующая система обозначений. Связная односторонняя кривая кодируется символом $\langle J\rangle$. Кривая, состоящая из одного овала, — символом $\langle 1\rangle$. Пустая кривая — символом $\langle 0\rangle$. Если символом $\langle A\rangle$ кодируется некоторый набор овалов, то кривая, получающаяся из него добавлением одного нового овала, охватывающего все старые, кодируется символом $\langle 1\langle A\rangle\rangle$. Кривая, представляющаяся в виде объединения двух непересекающихся кривых, которые кодируются символами $\langle A\rangle$ и $\langle B\rangle$ и таковы, что ни один овал одной из них не охватывается овалом другом, кодируется символом $\langle A \amalg B\rangle$. Кроме того, допускаются два сокращения: во-первых, если $\langle A\rangle$ — код некоторой кривой, то фрагмент другого кода, имеющий вид $A \amalg \ldots \amalg A$, где A повторяется n раз, сокращенно обозначается через $n\times A$ и, во-вторых, фрагменты кода, имеющие вид $n\times 1$, сокращенно обозначаются через n.

Эта система кодирования преобразуется в систему кодирования комплексных схем следующим образом. В зависимости от типа кривой весь код снабжается индексом I или II. В случае типа I коды положительных овалов снабжаются верхним индексом $^+$, а коды отрицательных — верхним индексом $^-$.

3. Ограничения на схемы кривых, известные к 1978 году

Ограничения на *вещественные* схемы кривых степени m, имеющие топологическое происхождение (подробности см. в [21] и [6]):

(3.1) $l\leqslant (m^2-3m+3+(-1)^m)/2$. Неравенство Харнака.

В следующих ограничениях (3.2)–(3.9) степень m предполагается четной, $m=2k$.

$$\left.\begin{array}{l}(3.2)\ \textit{Если } l=(m^2-3m+4)/2, \textit{ то } p-n\equiv k^2 \bmod 8.\\ (3.3)\ \textit{Если } l=(m^2-3m+2)/2, \textit{ то } p-n\equiv k^2\pm 1 \bmod 8.\end{array}\right\}\ \text{Экстремальные сравнения неравенства Харнака.}$$

$$\left.\begin{array}{l}(3.4)\ p-n^-\leqslant (3k^2-3k+2)/2.\\ (3.5)\ n-p^-\leqslant (3k^2-3k)/2.\end{array}\right\}\ \text{Усиленные неравенства Петровского.}$$

$$\left.\begin{array}{l}(3.6)\ p^-+p^0\leqslant (k^2-3k+3+(-1)^k)/2.\\ (3.7)\ n^-+n^0\leqslant (k^2-3k+2)/2.\end{array}\right\}\ \text{Усиленные неравенства Арнольда.}$$

(3.8) *Если k четно и $p^-+p^0=(k^2-3k+4)/2$, то $p^-=p^+=0$.*

(3.9) *Если k нечетно и $n^-+n^0=(k^2-3k+2)/2$, то $n^-=n^+=0$ и имеется только один внешний овал.*

Кроме неравенства Харнака имелось лишь одно ограничение топологического происхождения, относящееся не только к случаю четного m.

(3.10) *Если* $m \neq 4$, *то* $l^- + l^0 \leqslant (m-3)^2/4 + (m^2 - h^2)/4h^2$, где h — наибольшая степень простого числа, делящая m. В случае четного m оно следует из усиленных неравенств Арнольда и их экстремальных свойств; в случае нечетного m это неравенство Виро–Звонилова [21].

Теперь сформулирую ограничения топологического происхождения на комплексные схемы кривых степени m, обсуждавшиеся в [21].

(3.11) *Если кривая разбивающая, то* $l \equiv [m/2] \bmod 2$.

(3.12) *Если* $l = (m^2 - 3m + 3 + (-1)^m)/2$, *то кривая разбивающая.*

(3.13) *Если* m *четно и кривая разбивающая, то* $2(\Pi^+ - \Pi^-) = l - m^2/4$.

(3.14) *Если* m *неечтно и кривая разбивающая, то* $\Lambda^+ - \Lambda^- + 2(\Pi^+ - \Pi^-) = l - (m^2-1)/4$.

(3.15) *Если* m *четно,* $l = (m^2 - 3m)/2$ *и* $p - n \equiv m^2/4 + 4 \bmod 8$, *то кривая разбивающая.* Сравнение Харламова.

(3.16) *Если* $m \equiv 0 \bmod 4$ *и* $p^- + p^0 = (m^2 - 6m + 16)/8$, *то кривая разбивающая.*

(3.17) *Если* $m \equiv 2 \bmod 4$ *и* $n^- + n^0 = (m^2 - 6m + 8)/8$, *то кривая разбивающая.*

(3.16) и (3.17) являются экстремальными свойствами неравенств (3.6) и (3.7).

Ограничения, не имеющие топологического происхождения, как правило, труднообозримы. Прежде всего, к ним относятся следствия теоремы Безу, т.е. топологические следствия того факта, что неприводимые кривые степеней m и q либо совпадают либо пересекаются не более чем в mq точках. Сформулирую некоторые из них. Более общие формулировки см. в [9] и [21]. При $m \leqslant 11$ другие следствия теоремы Безу вытекают из формулируемых здесь.

(3.18) $h_2 \leqslant m/2$.

(3.19) *Если* $h_1 = [m/2]$, *то* $l = [m/2]$.

(3.20) $h_5' \leqslant m$.

(3.21) *Если* $h_4' = m$, *то* $l = m$.

(3.22) $h_8 \leqslant 3m/2$.

(3.23) *Если* $h_7 = [3m/2]$, *то* $l = [3m/2]$.

(3.24) $h_{13}' \leqslant 2m$.

(3.25) *Если* $h_{12}' = 2m$, *то* $l = 2m$.

Кроме ограничений этого типа в обзоре Рохлина [3] имеется только одно ограничение на комплексные схемы алгебраических кривых, не доказанное для схем гибких кривых:

(3.26) *Если* $h_1 = [m/2]$, *то кривая разбивающая.*

Рохлин [21, 3.6] заметил, что, модифицируя доказательство, можно получить и другие ограничения, однако он не обратил внимание на следующее ограничение, которое получается таким образом и не вытекает из перечисленных выше.

(3.27) *Если* $h_4' = m$, *то кривая разбивающая.*

4. Новые ограничения на схемы кривых

Рассмотрим сначала ограничения топологического происхождения.

Неравенство Рохлина [22]. (4.1) *Если кривая разбивающая и* $m \equiv 0 \bmod 4$, *то* $4\nu + p - n \leqslant (m^2 - 6m + 16)/2$.

(4.2) *Если кривая разбивающая и* $m \equiv 2 \bmod 4$, *то* $4\pi + n - p \leqslant (m^2 - 6m + 14)/2$.

Сопоставление этих теорем с (3.12) и (3.15) дает ограничения на *вещественные* схемы, доставляющие новую информацию при $m \geqslant 10$. Подробности см. в [22]. Недавно Фидлер (неопубликовано) нашел новые ограничения, аналогичные теоремам (4.1) и (4.2), которые доказываются по той же схеме, но с привлечением вспомогательных мнимых прямых и коник (последнее делает эти доказательства непригодными в случае гибких кривых).

В 1982 году независимо и разными средствами Т. Фидлер и В. В. Никулин [15] получили близкие ограничения. В обеих работах имеются результаты, относящиеся к кривым всех четных степеней, но при одних значениях степени сильнее результаты одной работы, при других — другой. Не обсуждая каждую работу в отдельности, сформулирую самые сильные результаты.

Теорема Фидлера (4.3). *Если* $m \equiv 4 \bmod 8$, $l = (m^2 - 3m + 4)/2$ *и характеристика каждого четного овала четна, то* $p - n \equiv -4 \bmod 16$.

Теоремы Никулина (4.4). *Если* $m \equiv 0 \bmod 8$, $l = (m^2 - 3m + 4)/2$ *и характеристика каждого четного овала делится на* 2^r, *то либо* $p - n \equiv 0 \bmod 2^{r+3}$, *либо* $p - n = 4^q \chi$, *где* $q \geqslant 2$ *и* $\chi \equiv 1 \bmod 2$.

(4.5) *Если* $m \equiv 2 \bmod 4$, $l = (m^2 - 3m + 4)/2$, *характеристика каждого нечетного овала делится на* 2^r, *число внешних овалов сравнимо с* 1 *по модулю* 2^r, *то* $p - n \equiv 1 \bmod 2^{r+3}$.

(4.6) *Если* $m \equiv 0 \bmod 4$, *кривая разбивающая и характеристика каждого четного овала четна, то* $p - n \equiv 0 \bmod 8$.

У Никулина [15] имеется, кроме того, следующее утверждение, которое, однако, как заметил А. Л. Слепян еще в 1980 г., непосредственно вытекает из формулы Рохлина (3.13).

(4.7) *Если m четно, кривая разбивающая, характеристика каждого нечетного овала четна, а число внешних овалов нечетно, то* $p-n \equiv \equiv m^2/4 \bmod 8$.

Теоремы (4.4)–(4.7) получены как следствия общих теорем о вещественных алгебраических поверхностях. Этими теоремами Никулина [15] исчерпываются ограничения на топологию вещественных алгебраических поверхностей, найденные за последние 5 лет. Из-за недостатка места мы не будем обсуждать их, см. [15].

Обратимся теперь к ограничениям нетопологического происхождения. В рассматриваемый период был открыт новый класс таких ограничений. Они еще менее обозримы, чем следствия теоремы Безу. Удовлетворительные общие формулировки для них не найдены. Поэтому ограничусь обсуждением источников, несколькими специальными формулировками и ссылкой на работы [23] и [5].

Большая часть новых ограничений доказывается построением вспомогательных кривых типа I (как правило, степеней 1 и 2) и применением теоремы Безу вместе с теоремами Фидлера о чередовании ориентаций или вместе со следующей теоремой (4.8). Теоремы о чередовании ориентаций слишком громоздки для изложения здесь (см. [23], а также [5]). В наиболее важных случаях они являются простыми следствиями теоремы (4.8), см. [5, 1.4].

(4.8) *Пусть A_1, A_2 — разбивающие кривые степеней m_1, m_2, трансверсальные друг другу и имеющие r вещественных точек пересечения, и пусть C — кривая степени $m = m_1+m_2$, получающаяся из $A_1 \cup A_2$ в результате малого возмущения, такая что некоторые комплексные ориентации кривых A_1, A_2 определяют ориентацию кривой $\mathbf{R}C$. Для $\mathbf{R}C$ с этой ориентацией положим*

$$\sigma = \begin{cases} m^2/4 - l + 2(\Pi^+ - \Pi^-), & \text{если } m \text{ четно},\\ (m^2-1)/4 - l + 2(\Pi^+ - \Pi^-) + \Lambda^+ - \Lambda^-, & \text{если } m \text{ нечетно}. \end{cases}$$

Тогда $0 \leqslant \sigma \leqslant m-r$; кривая C принадлежит типу I, *если и только если $\sigma = 0$, т.е. если и только если для ориентации ее вещественной части, определяемой комплексными ориентациями кривых A_1 и A_2, выполняется формула Рохлина; в этом случае эта ориентация является комплексной.*

Частные случаи этой теоремы были найдены Фидлером (см. [21, 3.7]), Марэном и Полотовским, окончательный вариант — Звониловым и автором. Она является частным случаем обобщения формул Рохлина на особые кривые, найденного Звониловым. Открытая как основа для построения кривых с заданной комплексной ориентацией, теорема (4.8) оказалась мощным ограничением на взаимные расположения двух кривых типа I, а при помощи вспомогательных построений её удалось превратить в источник ограничений на схемы индивидуальных кривых. Вот несколько таких ограничений.

(4.9) *Не существует кривой степени 7 со схемой* $\langle J \amalg 1\langle 14\rangle\rangle$.

(4.10) *Пусть* $\langle \alpha \amalg 1\langle\beta\rangle \amalg 1\langle\gamma\rangle \amalg 1\langle\delta\rangle\rangle$ — *вещественная схема кривой степени 8 с отличными от нуля* β, γ *и* δ. (i) *Если* $l = 22$ (*т.е.* $\alpha+\beta+\gamma+\delta = 19$), *то числа* β, γ *и* δ *нечетны.* (ii) *Если* $l = 20$ *и* $p - n \equiv 4 \bmod 8$, *то из чисел* β, γ *и* δ *два нечетны и одно четно.*

Теорема (4.9) была анонсирована в моей заметке [3] и приведена с разными доказательствами в [5] и [23]. Теорема (4.10) (i) при $\alpha = 0$, $\beta = 1$ была найдена Фидлером [23]. В полной общности теорема (4.10) доказана в [5]. Первые ограничения этого типа были найдены Фидлером. Идея воспользоваться вместо фидлеровского чередования ориентаций теоремой (4.8) принадлежит Рохлину.

Другое применение теоремы (4.8) и теорем о чередовании ориентаций — открытие связи между расположением в $\boldsymbol{R}P^2$ кривой относительно прямых и коник, с одной стороны, и комплексной схемы кривой, с другой стороны. Например, оказалось, что тип кривой степени 5 с $l = 4$ определяется ее расположением относительно прямых. Подробности см. в работе Фидлера [23]. Аналогичные связи типа кривой с другим ее алгебро-геометрическим инвариантом — числом вещественных θ-характеристик — открыты Гроссом и Харрисом [8].

5. Утверждение Клейна

Более ста лет назад Ф. Клейн [13, стр. 155] несколько туманно написал, что кривые типа I *не допускают развития.*

В 1978 году В. А. Рохлин [21, п. 3.9], ссылаясь на фактический материал и на эту фразу Клейна, высказал гипотезу, согласно которой любая кривая данной степени с данной вещественной схемой принадлежит типу I, если и только если эта схема не является частью большей вещественной схемы кривой той же степени.

Эта гипотеза не доказана и не опровергнута, но породила новые гипотезы. Как заметил Г. М. Полотовский, из ее справедливости и из теоремы (4.10) (i) вытекало бы следующее новое ограничение[1] на вещественные схемы кривых степени 8: если $\langle \alpha \amalg 1\langle\beta\rangle \amalg 1\langle\gamma\rangle \amalg 1\langle\delta\rangle\rangle$ —

[1] *Примечание при корректуре*: Гипотеза Рохлина частично опровергнута Е. И. Шустиным. Он построил, в частности, (M-2)-кривую со схемой $\langle 10 \amalg 1\langle 1\rangle \amalg 1\langle 2\rangle \amalg 1\langle 4\rangle\rangle$, не удовлетворяющей гипотезе Полотовского.

вещественная схема кривой степени 8, то при $20 \leqslant l \leqslant 21$ и $\beta+\gamma+\delta \equiv \equiv 3 \bmod 4$ числа β, γ и δ должны были бы быть нечетными, а при $19 \leqslant \leqslant l \leqslant 21$ и $\beta+\gamma+\delta \equiv 2 \bmod 4$ из этих чисел два должны бы быть нечётными, а одно — четным.

Мне кажется, что, несмотря на привлекательность и фундаментальность гипотезы Рохлина, слова Клейна нужно понимать более буквально. Именно, имеется следующая несложная теорема, которая по духу и по доказательству близка к пионерской работе Клейна:

(5.1) *Если A_t — непрерывное семейство вещественных алгебраических кривых (необязательно полоских), A_0 имеет только одну особую точку, и эта точка — невырожденная двойная, остальные A_t неособы, и если кривые A_t с $t<0$ разбивающие, то число компонент кривой $\boldsymbol{R}A_t$ с $t>0$ не превосходит числа компонент кривой $\boldsymbol{R}A_t$ с $t<0$.*

По-видимому, эта теорема является частным случаем теоремы о многообразиях произвольной размерности (ср. следующий пункт). Недавно А. Марэн (личное сообщение) переоткрыл теорему (5.1) и с ее помощью доказал, что *во всяком пучке плоских кривых четной степени $m \geqslant 4$ имеется кривая с числом компонент $\leqslant (m^2-3m-2)/2$* (последнее для $m \equiv 0 \bmod 4$ было опубликовано с неправильным доказательством А. Л. Чепониусом [28], правильное доказательство для $m=4$ было найдено Ю. С. Численко [30]). В случае $m=4$ отсюда следует, что через любые 13 вещественных точек проходит связная вещественная кривая степени 4, из чего очевидным образом вытекают теоремы (3.24) и (3.25). Вероятно, можно ожидать дальнейшего прогресса в топологическом исследовании пучков вещественных кривых, которое интересно и само по себе и как источник ограничений нетопологического происхождения на схемы кривых.

6. Комплексные топологические характеристики поверхностей

Перенесение теории комплексных топологических характеристик кривых на случай больших размерностей только начинается, и делать обзор еще рано. Сформулирую лишь несколько определений и фактов, относящихся к случаю поверхностей.

Имеются три типа неособых вещественных алгебраических поверхностей: $\mathrm{I}_{\mathrm{abs}}$, $\mathrm{I}_{\mathrm{rel}}$ и II. Поверхность A относится: к типу $\mathrm{I}_{\mathrm{abs}}$, если ее вещественная часть $\boldsymbol{R}A$ реализует нуль группы $H_2(\boldsymbol{C}A; \boldsymbol{Z}_2)$; к типу $\mathrm{I}_{\mathrm{rel}}$, если $\boldsymbol{R}A$ и плоские сечения поверхности $\boldsymbol{C}A$ реализуют один элемент группы $H_2(\boldsymbol{C}A; \boldsymbol{Z}_2)$; к типу II в остальных случаях.

Для поверхностей типов $\mathrm{I_{abs}}$ и $\mathrm{I_{rel}}$ автор [4] определил структуры, аналогичные комплексным ориентациям кривых типа I. Это определение удобно сформулировать в аналитической ситуации, обобщающей случаи поверхностей типов $\mathrm{I_{abs}}$ и $\mathrm{I_{rel}}$. Пусть $\mathscr{X}$ — неособая комплексная поверхность с $H_1(\mathscr{X}; \boldsymbol{Z}_2) = 0$, пусть $c\colon \mathscr{X} \to \mathscr{X}$ — антиголоморфная инволюция и $\mathscr{Y} \subset \mathscr{X}$ — неособая кривая (возможно, $\mathscr{Y} = \varnothing$), инвариантная относительно c. Положим $X = \mathrm{fix}(c)$ и $Y = X \cap \mathscr{Y}$. Пусть X и $\mathscr{Y}$ реализуют один элемент группы $H_2(\mathscr{X}; \boldsymbol{Z}_2)$ (в случае $\mathscr{Y} = \varnothing$ это означает, что X реализует нуль). Тогда $X \setminus Y$ обладает двумя выделенными противоположными ориентациями и выделенной спинорной структурой, которые называются *комплексными* и определяются следующими свойствами.

Пусть a_0 и a_1 — точки из $X \setminus Y$. Комплексная ориентация поверхности $X \setminus Y$ и естественная ориентация многообразия $\mathscr{X}$ определяют ориентации слоев D_0 и D_1 трубчатой окрестности поверхности X в $\mathscr{X}$, лежащих над a_0 и a_1. Пусть $b_i \in \partial D_i$ и u_i — путь в ∂D_i, соединяющий b_i с $c(b_i)$ и согласованный с ориентацией диска D_i, и пусть $s\colon I \to \mathscr{X} \setminus (X \cup \mathscr{Y})$ — путь, соединяющий b_0 с b_1. Тогда петля $su_1(c \circ s)^{-1}u_0^{-1}$ реализует нуль группы $H_1(\mathscr{X} \setminus (X \cup \mathscr{Y}); \boldsymbol{Z}_2)$.

Квадратичная форма $H_1(X \setminus Y; \boldsymbol{Z}_2) \to \boldsymbol{Z}_2$, отвечающая (см. [12]) выделенной спинорной структуре, относит классу, реализуемому гладко вложенной окружностью S, измененный на 1 коэффициент зацепления в $\mathscr{X}$ объединения $X \cup \mathscr{Y}$ и окружности, которая получается из S в результате сдвига вдоль касательного к S векторного поля, умноженного на $\sqrt{-1}$.

Комплексные ориентации и спинорные структуры поверхностей допускают применения, аналогичные применениям комплексных ориентаций кривых. Сформулирую только одну теорему, доказанную с их помощью — обобщение утверждения Клейна на случай поверхностей, доказанное В. М. Харламовым и автором (не опубликовано). *Если A_t — непрерывное семейство вещественных алгебраических поверхностей трехмерного проективного пространства, A_0 имеет только одну особую точку, и эта точка — невырожденная двойная, остальные A_t неособы, и если A_t с $t < 0$ принадлежат типу $\mathrm{I_{abs}}$, то число $\dim H_*(\boldsymbol{R}A_t; \boldsymbol{Z}_2)$ с $t > 0$ не превосходит $\dim H_*(\boldsymbol{R}A_t; \boldsymbol{Z}_2)$ с $t < 0$.*

7. Изотопии

Напомню, что *жесткой изотопией* кривых степени m называется изотопия в классе (неособых) кривых степени m (т.е. путь в пространстве кривых степени m). Это понятие естественно вводится и для других аналогичных классов многообразий; например, для поверхностей в $\mathbb{R}P^3$.

Классификация кривых степени $\leqslant 4$ и поверхностей степени $\leqslant 3$ относительно жестких изотопий была известна в прошлом веке. С точностью до жестких изотопий кривая степени $\leqslant 4$ определяется своей вещественной схемой, а поверхность степени $\leqslant 3$ — топологическим типом своей вещественной части.

Классификация кривых степеней 5 и 6 относительно жестких изотопий была завершена в 1978–1980 годах в работах Рохлина [21], Никулина [14] и Харламова [26]. Рохлин [21] показал, что для кривых степени $\geqslant 5$ класс кривой относительно жестких изотопий уже не определяется вещественной схемой, поставил вопрос, до какой степени он определяется комплексной схемой и дал классификацию комплексных схем кривых степени $\leqslant 6$, показав, в частности, что комплексная схема кривой степени $\leqslant 6$ определяется вещественной схемой и типом кривой. То, что вещественная схема и тип действительно определяют кривую с точностью до жестких изотопий, было доказано для кривых степени 5 — Харламовым [26] и для кривых степени 6 — Никулиным [14].

Уже для кривых степени 7 проблема классификации относительно жестких изотопий не решена. Имеются примеры, показывающие, что ни вещественная схема и тип, ни даже комплексная схема не определяют кривую степени 7 с точностью до жестких изотопий. Еще в работе Рохлина [21] были построены кривые степени 7 типа I с вещественной схемой $\langle J \amalg 3 \amalg 1\langle 3\rangle\rangle$, различающиеся комплексными ориентациями, и потому не являющиеся жестко изотопными (их комплексные схемы — $\langle J \amalg 1^+ \amalg 2^- \amalg 1^- \langle 1^+ \amalg 2^-\rangle\rangle_{\mathrm{I}}$ и $\langle J \amalg 3^- \amalg 1^+ \langle 3^+\rangle\rangle_{\mathrm{I}}$). Марэн [18] и Фидлер [23] заметили, что если вещественная схема кривой степени m обуславливает существование прямой, пересекающейся с этой кривой в m вещественных точках (т.е. если неравенство (3.18) обращается в равенство), то расположение овалов кривой относительно такой прямой сохраняется в процессе жесткой изотопии. Фидлер [23] построил основанные на этом наблюдении примеры кривых степени 7 типа II (в частности, кривые с той же схемой $\langle J \amalg 3 \amalg 1\langle 3\rangle\rangle$), не являющиеся жестко изотопными. (Марэн построил аналогичны примеры с другой вещественной схемой.)

Гибкой изотопией назовём изотопию в классе гибких кривых. Так как комплексная схема сохраняется при гибкой изотопии, то для кривых степени $\leqslant 6$ гибкая изотопность равносильна жесткой. Рохлин

[21] выдвинул предположение, что кривые одной степени с одной комплексной схемой гибко изотопны (в [21] гибкие изотопии называются эквивариантными изотопиями). Это предположение и примеры Фидлера [23] дают основание предполагать, что для кривых степени 7 гибкая изотопность уже не равносильна жесткой. Дополнительным по отношению к комплексной схеме инвариантом относительно гибких изотопий мог бы оказаться гомотопический тип пространства $CP^2\setminus(\boldsymbol{R}P^2\cup CA)$ с инволюцией conj. Первое его обследование выполнено С. М. Финашиным [24]: для большого класса кривых (в частности, для кривых степени $\leqslant 5$) вычислена фундаментальная группа этого пространства, и построены две кривые степени 7 с комплексной схемой $\langle J \amalg 2 \amalg 1\langle 2\rangle\rangle_{II}$, которые имеют гомотопически эквивалентные пространства $CP^2\setminus(\boldsymbol{R}P^2\cup CA)$, но не являются жестко изотопными. Весьма вероятно, что они гибко изотопны.

Проблема классификации поверхностей степени 4 относительно жестких изотопий оказались тоньше, чем соответствующие проблемы для кривых степеней 5 и 6. Сначала была решена более грубая проблема. Было доказано, что поверхность степени 4 определяется с точностью до жестких изотопий и проективных преобразований своим типом (см. § 7), топологией вещественной части и ее расположением в $\boldsymbol{R}P^3$. См. Никулин [14].

Недавно Харламов показал, что некоторые поверхности степени 4 не являются жестко изотопными своим зеркальным образам (такова, в частности, поверхность, гомеоморфная несвязной сумме сферы с двумя ручками и девяти сфер), и редуцировал проблему классификации поверхностей степени 4 относительно жестких изотопий к задаче об арифметических группах.[2]

9. Построения

В классических работах топологии вещественных алгебраических кривых построения производились следующим образом: сначала строились две неособые трансверсальные друг другу кривые, затем их объединение подвергалось малому возмущению, устраняющему особенности. Для реализации двух вещественных схем кривыми степени 6 Д. А. Гудкову пришлось выйти за рамки этой схемы и возмущению подвергать не распадающуюся кривую, а образ неособой кривой при

[2] Примечание при корректуре: Осенью 1983 г. Харламов завершил классификацию поверхностей степени 4 относительно жестких изотопий.

квадратичном преобразовании. Однако, как и прежде, все возмущаемые кривые имели только невырожденные двойные особенности. Появлению в контексте топологии вещественных алгебраических кривых возмущений сложных особенностей мешали, по-видимому, два обстоятельства: во-первых, не очень сложные особенности не дают ничего нового по сравнению с невырожденными двойными точками (выгода от усложнения особенностей, как выяснилось, появляется лишь при переходе к невырожденным пятикратным точкам и точкам касания трех ветвей); во-вторых, для целенаправленного возмущения кривых со сложными особенностями необходимо было разработать специальную технику.

В 1980 году я предложил конструкцию, которая позволяет строить возмущения кривой с полуквазиоднородной особенностью, устраняющие эту особенность и заменяющие некоторую окрестность особой точки заранее заготовленным фрагментом. Для некоторых полуквазиоднородных особенностей (например, для точек невырожденного касания трех ветвей) мне удалось получить полную топологическую классификацию их устранений, т.е. удалось расклассифицировать с точностью до топологической эквивалентности фрагменты кривой, появляющиеся на месте особенности в результате ее устранения. Для некоторых других особенностей (например, для невырожденных пятикратных точек и точек невырожденного касания четырёх ветвей) удалось сконструировать обширный запас их устранений. Ю. С. Численко [29] продолжила эту работу и сконструировала много устранений для точек невырожденного касания пяти ветвей.

Новый метод построения, который представляет собой соединение конструкции устранения полуквазиоднородных особенностей с некоторыми старыми приемами построений и с широким применением кремоновых преобразований, позволил решить несколько задач, недоступных для старых методов. Сформулирую основные результаты, полученные с его помощью (см. [3]).

Прежде всего, мне удалось завершить классификацию вещественных схем кривых степени 7.

(9.1) *Существуют кривые степени* 7 *со следующими вещественными схемами*

(i) $\langle J \amalg \alpha \amalg 1\langle\beta\rangle\rangle$ *с* $\alpha+\beta\leqslant 14,\ 0\leqslant\alpha\leqslant 13,\ 1\leqslant\beta\leqslant 13$;

(ii) $\langle J \amalg \alpha\rangle$ *с* $0\leqslant\alpha\leqslant 15$;

(iii) $\langle J \amalg 1\langle 1\langle 1\rangle\rangle\rangle$.

Любая кривая степени 7 *имеет одну из этих* 121 *вещественных схем.*

К 1980 году оставалось неизвестным, существуют ли кривые типов $\langle J \amalg 1\langle 14\rangle\rangle$, $\langle J \amalg 10 \amalg 1\langle 4\rangle\rangle$ и $\langle J \amalg \alpha \amalg 1\langle\beta\rangle\rangle$ с $13 \leqslant \alpha+\beta \leqslant 14$, $3 \leqslant \alpha$, $6 \leqslant \beta$. Нереализуемость схемы $\langle J \amalg 1\langle 14\rangle\rangle$ есть содержание сформулированной выше теоремы (5.2). Схемы $\langle J \amalg \alpha \amalg 1\langle\beta\rangle\rangle$ с $6 \leqslant \alpha+\beta \leqslant 14$, $1 \leqslant \alpha$, $2 \leqslant \beta$ реализуются следующим образом: строятся 4 кривые степени 7, имеющие по две особые точки, в каждой из которых друг друга невырожденно касаются три неособые ветви, затем эти кривые возмущаются при помощи техники, о которой шла речь выше. Схема $\langle J \amalg 4 \amalg 1\langle 10\rangle\rangle$ не только не была, но и не могла быть реализована классическим методом, как заметили Звонилов и Фидлер.

Далее, удалось построить контрпримеры к одному из неравенств, составляющих гипотезу Рэгсдейл. Напомню, что эта гипотеза была сформулирована в 1906 году [18] и состоит в том, что для любой кривой четной степени $p \leqslant (3m^2-6m+8)/8$ и $n \leqslant (3m^2-6m)/8$. В 1938 году И. Г. Петровский [16] сформулировал независимо от Рэгсдейл более слабую гипотезу:

$$p \leqslant (3m^2-6m+8)/8 \quad \text{и} \quad n \leqslant (3m^2-6m+8)/8.$$

Мне удалось для любого $m \geqslant 8$, кратного 4, построить кривую степени m с вещественной схемой $\langle (m^2-6m)/8 \amalg 1\langle (3m^2-6m+8)/8\rangle\rangle$ и, тем самым, показать, что правое неравенство Рэгсдейл нарушается при $m \geqslant 8$, $m \equiv 0 \bmod 4$. Вопрос о справедливости гипотезы Петровского остается открытым. Он допускает более широкую формулировку: верно ли, что если X — множество неподвижных точек антиголоморфной инволюции неособой односвязной компактной комплексной поверхности $\mathscr{X}$, то $\dim H_1(X; \mathbf{Z}_2) \leqslant h^{1,1}(\mathscr{X})$.

Кроме того, при помощи нового метода были реализованы 42 новые вещественные схемы кривых степени 8 с $l = 22$, см. [3] (старые методы дали 10 схем) и около 500 новых вещественных схем кривых степени 10 с $l = 37$, см. [29] (старые методы дали 38 схем).

Вероятно, можно ожидать, что новый метод будет применен и для построения поверхностей. К описанным в обзоре Харламова построениям поверхностей можно добавить только построения из моей работы [2]. В ней строятся поверхности четной степени способом, состоящим в малом возмущении квадрата уравнения неособой поверхности в два раза меньшей степени. Построены поверхности больших степеней, опровергающие имевшуюся гипотезу о максимальном числе компонент поверхности заданной степени. Реализованы все изотопические типы поверхностей степени 4, кроме одного (прежние их реализации были менее элементарны — см. [25]).

Литература

[1] Арнольд В. И., О расположении овалов вещественных плоских алгебраических кривых, инволюциях четырехмерных гладких многообразий и арифметике целочисленных квадратичных форм, *Функц. анализ* **5** (3) (1971), стр. 1–9.

[2] Виро О. Я., Построение многокомпонентных вещественных алгебраических поверхностей, *Доклады АН СССР* **248** (2) (1979), стр. 279–282.

[3] Виро О. Я., Кривые степени 7, кривые степени 8 и гипотеза Рэгсдейл, *Доклады АН СССР* **254** (6) (1980), стр. 1305–1310.

[4] Виро О. Я., Комплексные ориентации вещественных алгебраических поверхностей, *УМН* **37** (4) (1982), стр. 96.

[5] Виро О. Я., Плоские вещественные кривые степеней 7 и 8: новые запреты, *Известия АН СССР, сер. матем.* **47** (5) (1983), стр. 1135–1150.

[6] Wilson G., Hilbert's Sixteenth Problem, *Topology* **17** (1978), pp. 53–74.

[7] Hilbert D., Mathematische Probleme, *Arch. Math. Phys.* **3** (1) (1901), pp. 213–237.

[8] Gross B. H. and Harris J., Real Algebraic Curves, *Ann. Sci. Ec. Norm. Sup. 4e serie* **14** (1981), pp. 157–182.

[9] Гудков Д. А., Топология вещественных проективных алгебраических многообразий, *УМН* **29** (4) (1974), стр. 3–79.

[10] Гудков Д. А., О топологии алгебраических кривых на гиперболоиде, *УМН* **34** (6) (1979), стр. 26–32.

[11] Звонилов В. И., Комплексные топологические характеристики вещественных алгебраических кривых на поверхности, *Функц. анализ* **16** (3) (1982), стр. 56–57.

[12] Johnson D., Spin-Structures and Quadratic Forms on Surfaces, *J. London Math. Soc.* **22** (2) (1980), pp. 365–373.

[13] Klein F., *Gesammelte mathematische Abhandlungen*, B. 2, Berlin, 1922.

[14] Никулин В. В., Целочисленные квадратичные формы и некоторые их геометрические приложения, *Известия АН СССР, сер. матем.* **43** (1) (1979), стр. 111–177.

[15] Никулин В. В., Инволюции целочисленных квадратичных форм и их приложения к вещественной алгебраической геометрии, *Известия АН СССР, сер. матем.* **47** (1) (1983), стр. 109–188.

[16] Petrovsky J. G., On the Topology of Real Plane Algebraic Curves, *Ann. Math.* **39** (1) (1938), pp. 187–209.

[17] Petrovsky J. G., On the Diffusion of Waves and the Lacunes for Hyperbolic Equations, *Матем. сб.* **17** (59) (1945), стр. 289-370.

[18] Ragsdale V., On the Arrangement of the Real Branches of Plane Algebraic Curves, *Amer. J. Math.* **28** (1906), pp. 377–404.

[19] Рохлин В. А., Сравнения по модулю 16 в шестнадцатой проблеме Гильберта, *Функц. анализ* **5** (4) (1972), стр. 58–64.

[20] Рохлин В. А., Комплексные ориентации вещественных алгебраических кривых, *Функц. анализ* **8** (4) (1974), стр. 71–75.

[21] Рохлин В. А., Комплексные топологические характеристики вещественных алгебраических кривых, *УМН* **33** (5) (1978), стр. 77–89.

[22] Рохлин В. А., Новые неравенства в топологии вещественных плоских алгебраических кривых, *Функц. анализ* **14** (1) (1980), стр. 37–43.

[23] Фидлер Т., Пучки прямых и топология вещественных алгебраических кривых, *Известия АН СССР, сер. матем.* **46** (4) (1982), стр. 853–863.

[24] Финашин С. М., Топология дополнения вещественной алгебраической кривой в CP^2, *Зап. науч. семинаров ЛОМИ* **122** (1982), стр. 137–145.

[25] Харламов В. М., Вещественные алгебраических поверхности, *Труды Международного конгресса математиков, Хельсинки*, **1978**, стр. 421–428.

[26] Харламов В. М., Жесткая изотопическая классификация вещественных плоских кривых степени 5, *Функц. анализ* **15** (1) (1981), стр. 88–89.

[27] Хованский А. Г., Индекс полиномиального векторного поля, *Функц. анализ* **13** (1) (1979), стр. 49–58.

[28] Чепониус А. Л., О гнездах вещественных плоских алгебраических кривых, *Литовский матем. сб.* **16** (4) (1976), стр. 239–243.

[29] Численко Ю. С., *M*-кривые десятой степени, *Зап. науч. семинаров ЛОМИ* **122** (1982), стр. 146–161.

[30] Численко Ю. С., Пучки вещественных алгебраических кривых, *Тезисы Ленинградской международной топологической конференции, Ленинград*, 1982, стр. 28.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

F. R. COHEN*

Applications of Loop Spaces to Classical Homotopy Theory

A simply-connected space X has exponent p^k at the prime p if p^k is the smallest positive integer which annihilates the p-primary component of $\pi_q X$ for all q. The first interesting example of a space with an exponent is the odd dimensional sphere S^{2n+1}. In the middle 1950's I. M. James proved that at $p = 2$, S^{2n+1} has an exponent which divides 2^{2n} [12]. H. Toda subsequently showed that if p is an odd prime then at p, S^{2n+1} has an exponent which divides p^{2n} [23]. In this note we shall summarize recent related work.

M. G. Barratt conjectured that S^{2n+1} has exponent p^n if p is an odd prime. The crux here of course is to show that p^n annihilates the p-primary component of $\pi_q S^{2n+1}$ for all q. B. Gray has exhibited elements of order p^n in the homotopy groups of S^{2n+1} [10].

In 1978 P. Selick proved that S^3 has exponent p if p is odd [22]. Using apparently different techniques the author, J. C. Moore, and J. A. Neisendorfer gave a proof of Barratt's conjecture in case p is a prime greater than 3 [8, 9]. Neisendorfer gave a proof in case $p = 3$ [17].

THEOREM 1. *If $p > 2$, S^{2n+1} has exponent p^n.*

The current status in case $p = 2$ is different. M. G. Barratt and M. Mahowald have made the following

CONJECTURE. S^{2n+1} has exponent $2^{\varphi(2n)}$ at the prime 2 where $2^{\varphi(2n)}$ is the order of the canonical line bundle over $\boldsymbol{R}P^{2n}$.

Mahowald has exhibited elements of order $2^{\varphi(2n)}$ in the homotopy groups of S^{2n+1} [14]. Small improvements in James' results were given

* The author was partially supported by the National Science Foundation and the Alfred P. Sloan Foundation.

by the author in [3] where this conjecture was checked for S^5 and within a factor of 4 for S^9. P. Selick then showed that the exponent of S^{4n+1} divides twice the exponent of S^{4n-1} [21]. However, the work in [8, 9, 17, 18] does not appear to generalize to the 2-primary case.

Our work in [9] leading to the proof of Theorem 1 initially seemed to have little to do with the homotopy groups of S^{2n+1}. We studied the Moore spaces $P^n(k) = S^{n-1} \cup_k e^n$ and wondered whether $P^n(k)$ had any exponent at all if $k > 1$. In case k is odd, we obtained the following theorem [6].

THEOREM 2. *If $p > 2$ and $n > 2$, then $P^n(p^r)$ has an exponent which divides p^{2r+1}.*

Notice that the Hurewicz isomorphism theorem implies that $\pi_{n-1}P^n(p^r)$ is isomorphic to $Z/p^r Z$. We showed that there are infinitely many elements of order p^{r+1} in the homotopy groups of $P^n(p^r)$ [9]. For example, we have

THEOREM 3. *If $p > 2$ and $n > 1$, then $\pi_{2np^k-1} P^{2n+1}(p^r)$ has $Z/p^{r+1}Z$ summand for all $k \geqslant 1$.*

About one year before Theorem 2 was proven M. G. Barratt made the following

CONJECTURE. If X is a pointed space such that the suspension order of the identity of $\Sigma^2 X$ is p^n, then the identity map of $\Omega^2\Sigma^2 X$ has order p^{n+1}.

Since the p^{n+1}-st power map on an H-space induces multiplication by p^{n+1} on homotopy groups, Barratt's "finite exponent conjecture" above implies that $P^n(p^r)$ has exponent p^{r+1} if $n > 3$ and $p > 2$. Refining the techniques in the proof of Theorem 2, Neisendorfer gave the first and essentially only known non-trivial example of this last conjecture [16].

THEOREM 4. *If $p > 2$ and $n > 2$, then $P^n(p^r)$ has exponent p^{r+1}.*

It is not yet known whether $P^n(2^r)$ has an exponent if $n > 2$. The case of $P^2(k)$ is straightforward since the universal cover of $P^2(k)$ is homotopy equivalent to a $(k-1)$-fold bouquet of 2-spheres. Furthermore, the bouquet $S^j \vee S^n$, $j, n \geqslant 2$, does not have exponent at all since, by the Hilton–Milnor theorem [11, 15], the loop space of $S^j \vee S^n$ is homotopy equivalent to a weak product of loop spaces of arbitrarily large spheres.

The proofs of Theorems 1, 2, and 4 yielded more information than these results actually state. We give some of this additional information. Let $\Omega^n X$ denote the n-fold based loop space of X. In the next two theorems [8, 16, 22] assume that all spaces are localized at the prime p.

THEOREM 5. *If $p > 2$, the identity map of the universal cover of $\Omega^{2n}S^{2n+1}$ has order p^n.*

THEOREM 6. *If $p > 2$ and $n > 2$, the identity map of $\Omega^2 P^n(p^r)$ has order p^{r+1}.*

Barratt in the late 1950's obtained bounds in the growth of the order of the torsion in the homotopy groups of ΣX if the suspension order of the identity is finite [1]. A proof which is somewhat different is given in [4]. However, there has been little progress on the "finite exponent conjecture" in general.

A striking further conjecture was made by J. C. Moore:

CONJECTURE. Assume that X is a simply-connected finite complex. X has an exponent (at every prime p) if and only if $\pi_* X \otimes Q$ is a totally finite dimensional vector space.

For example, $S^j \vee S^n, j, n \geqslant 2$, does not have an exponent and $\pi_*(S^j \vee S^n) \otimes Q$ is infinite dimensional. Neisendorfer and Selick showed that the 2-cell complex $Y = S^j \cup_\alpha e^n$ does not have an exponent if Y has been localized at an odd prime and α has finite order [19]. Using different techniques [4], the author gave a proof of a similar result in case $p = 2$ and α is a double suspension. Selick then showed that if X is a suspension localized at an odd prime with torsion free homology, then the condition that $\pi_* X \otimes Q$ is infinite dimensional implies that X does not have an exponent [20]. It seems likely that Selick's techniques should apply in case $p = 2$. Unfortunately, there are few positive results which exhibit spaces with exponents.

The heart of the proofs of Theorems 1, 2, 4, 5 and 6 is the fact that $\Omega P^n(p^r)$, $n > 2$, and related loop spaces decompose as a non-trivial product in the homotopy category. Sufficiently tight control of the "pieces" yields the requisite proofs. A survey of these decompositions together with related techniques is given in [7] to which the interested reader is strongly referred. Although $\Omega\Sigma^2 X$ decomposes into a non-trivial product if $\Sigma^2 X$ is a p-local complex which is not homotopy equivalent to a sphere, the "pieces" are far from being well understood; these general decompositions shed very little light on general exponent questions as yet.

We close this note by relating P. Selick's results on S^3 and the Kahn–Priddy theorem [13]. Assume that all spaces are localized at the prime p. Let $g\colon \Omega P^{2p+2}(p) \to S^3$ be any map which induces an isomorphism on π_{2p}. If $p > 2$, then g induces a split epimorphism on homotopy groups [5].

Selick's original work applied to a choice of g obtained by looping the natural map $P^{2p+2}(p)\to BS^3$ which induces an isomorphism on π_{2p+1} [22]. This map fits nicely into a more general setting.

Whitehead [24] defines maps $\Sigma^k \mathbf{R}P^{k-1}\to S^k$ and Boardman and Vogt [2, p. 64] define maps Φ_k: $\Sigma^k B(\mathbf{R}^k, p)\to S^k$ where $B(\mathbf{R}^k, p)$ is the configuration space of unordered p-tuples of distinct points in $\mathbf{R}^k$. In case $p=2$, $B(\mathbf{R}^k, 2)$ has the homotopy type of $\mathbf{R}P^{k-1}$ and in case $k=\infty$, $B(\mathbf{R}^\infty, p)$ is a $K(\Sigma_p, 1)$ where Σ_p is the symmetric group on p letters. If in addition S^k is a (p-local) H-space, Φ_k extends to a map θ_k: $\Omega\Sigma^{k+1}B(\mathbf{R}^k, p)\to S^k$. Thus if $p>2$, there is a homotopy commutative diagram

$$\begin{tikzcd}
\Omega^4\Sigma^4 B(\mathbf{R}^3, p) \arrow[r] \arrow[d, "\Omega^3\theta_3"] & \dots \arrow[r] & \Omega^{2n+2}\Sigma^{2n+2}B(\mathbf{R}^{2n+1}, p) \arrow[r] \arrow[d, "\Omega^{2n+1}\theta_{2n+1}"] & \dots \arrow[r] & \Omega^\infty\Sigma^\infty K(\Sigma_p, 1) \arrow[d] \\
\Omega^3 S^3 \arrow[r] & \dots \arrow[r] & \Omega^{2n+1}S^{2n+1} \arrow[r] & \dots \arrow[r] & \Omega^\infty S^\infty .
\end{tikzcd}$$

Since $\Omega\Sigma^4 B(\mathbf{R}^3, p)$ is homotopy equivalent to $\Omega P^{2p+2}(p)$ and θ_3 induces an isomorphism on π_{2p}, $\Omega^3\theta_3$ induces a split epimorphism on π_q, $q>0$. That the right-hand vertical map induces a split epimorphism on π_q, $q>0$, is a restatement of the Kahn–Priddy theorem [13]. The following conjecture is made in [3].

CONJECTURE. If $p>2$, $\Omega^{2n+1}\theta_{2n+1}$ induces a split epimorphism on π_q, $q>0$. If $p=2$, φ_{2n+1} induces a (non-split) epimorphism on π_q, $q>0$.

The relationship between this conjecture and exponents is as follows. If $p>2$, the suspension order of the identity for $\Sigma^{2n+1}B(\mathbf{R}^{2n+1}, p)$ is precisely p^n and S^{2n+1} supports a homotopy abelian H-space structure. Thus the image on homotopy groups induced by θ_{2n+1} is annihilated by p^n and so Theorem 1 would follow from the conjecture. Similarly, if $p=2$, the image on homotopy groups induced by φ_{2n+1} is annihilated by $2^{\varphi(2n)+1}$ which is a factor of two away from the conjectured value.

An important related point here is that self-maps f of certain loop spaces can be shown to be homotopy equivalences provided the map f satisfies certain minimal hypotheses. Some examples follow: Localize all spaces at $p=2$. If $n>1$ and f is a degree one self-map of $\Omega^2 S^{2n+1}$, then f is a homotopy equivalence [5]. If $n<\infty$, and f is a self-map of $Q\mathbf{R}P^{2n}$, $Q\,CP^n$, or the base-point component of QS^0 such that f is an H-map which induces an isomorphism on the first non-vanishing homotopy group, then f is a homotopy equivalence. Notice that any degree one self-map of the universal cover of QS^1 is a homotopy equivalence and so the Kahn–Priddy theorem follows. These last results will appear in work of the author, E. Campbell, F. Peterson, and P. Selick.

Notes added in proof: (1) C. McGibbon and C. Wilkerson have jointly obtained many examples of Moore's conjecture. They have proven that if X is a finite simply-connected CW complex with $(\pi_* X)\otimes Q$ a finite-dimensional rational vector space, there exist integers $A(X)$ and $B(X)$ such that for any prime $p > A(X)$, $p^{B(X)}\big(\mathrm{Torsion}(\pi_* X\otimes Z_{(p)})\big) = 0$. In the proof of their theorem, p must be quite large. In particular the integral homology of X has no p-torsion.

(2) The author has obtained a 2-primary analogue of Selick's theorem and the Kahn–Priddy theorem for the 3-sphere. Let $(\Omega X)\{2\}$ denote the homotopy theoretic fibre of the H-space squaring map for ΩX. There is a map Θ: $(\Omega^3 S^5)\{2\}\to\Omega_0^3 S^3$ which induces a split epimorphism on the 2-primary components of homotopy groups.

References

[1] Barratt M. G., Spaces of Finite Characteristic, *Quart. J. Math.* **11** (1960), pp. 124–136.

[2] Boardman J. M. and Vogt R. M., *Homotopy Invariant Algebraic Structures on Topological Spaces*, Springer-Verlag Lecture Notes in Math. 347, 1973.

[3] Cohen F. R., The unstable decomposition of $\Omega^2\Sigma^2 X$, *Math. Zeit.* **182** (1983), pp. 553–568.

[4] Cohen F. R., Orders of Certain Compositions, Current Trends in Algebraic Topology, *Canadian Math. Soc. Conference Proceedings* **2** (1981), Part 1, pp. 289–296.

[5] Cohen F. R. and Mahowald M., A Remark on the Self-Maps of $\Omega^2 S^{2n+1}$, *Indiana Univ. Math. J.* **4** (1981), pp. 583–588.

[6] Cohen F. R., Moore J. C. and Neisendorfer J. A., *Exponents in Homotopy Theory*, to appear.

[7] Cohen F. R., Moore J. C., and Neisendorfer J. A., *Decomposition of Loop Spaces and Applications to Exponents*, Algebraic Topology Aarhus 1978, Springer-Verlag Lecture Notes in Math. 763 (1978), pp. 1–11.

[8] Cohen F. R., Moore J. C. and Neisendorfer J. A., The Double Suspension and p-Primary Components of the Homotopy Groups of Spheres, *Ann. of Math.* **110** (1979), pp. 549–565.

[9] Cohen F. R., Moore J. C. and Neisendorfer J. A., Torsion in Homotopy Groups, *Ann. of Math.* **109** (1979), pp. 121–168.

[10] Gray B., On the Sphere of Origin of Infinite Families in the Homotopy Groups of Spheres, *Topology* **8** (1969), pp. 219–232.

[11] Hilton P. J., On the Homotopy Groups of a Union of Spheres, *Jour. London Math. Soc.* **30** (1955), pp. 154–172.

[12] James I. M., On the Suspension Sequence, *Ann. of Math.* **65** (1957), pp. 74–107.
[13] Kahn D. S. and Priddy S. B., The Transfer and Stable Homotopy Theory, *Math. Proc. Cambridge Phil. Soc.* **83** (1978), pp. 103–111.
[14] Mahowald M., The Image of J in the EHP Sequence, *Ann. of Math.* **116** (1982), pp. 65–112.
[15] Milnor J., *On the Construction* FK, in "*Algebraic Topology – A Student's Guide*" by J. F. Adams, Cambridge Univ. Press, 1972.
[16] Neisendorfer J. A., *The Exponent of a Moore Space*, to appear.
[17] Neisendorfer J. A., 3 Primary Exponents, *Math. Proc. Cambridge Phil. Soc.* **90** (1981), pp. 163–183.
[18] Neisendorfer J. A., Primary Homotopy Theory, *Memoirs of the A. M. S.*, **232**, 1980.
[19] Neisendorfer J. A. and P. Selick, Some Examples of Spaces With and Without Exponents, Current Trends in Algebraic Topology, *Canadian Math. Soc. Conference Proceedings* **2** (1981), part 1, pp. 343–358.
[20] Selick P., On Conjectures of Moore and Serre, to appear in *Math. Proc. Cambridge Phil. Soc.*.
[21] Selick P., *On the 2-Primary Exponents of the Homotopy Groups of Spheres*, to appear.
[22] Selick P., Odd Primary Torsion in $\pi_k(S^3)$, *Topology* **17** (1978), pp. 407–412.
[23] Toda H., On the Double Suspension E^2, *J. Inst. Polytech., Osaka City Univ., Ser.* A, **7** (1956), pp. 103–145.
[24] Whitehead G. W., *Recent Advances in Homotopy Theory*, Regional Conference Series in Math., no. 5, American Math. Soc., Providence, R. I., 1970.

UNIVERSITY OF KENTUCKY
LEXINGTON KENTUCKY 40506
USA

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

RALPH L. COHEN*

The Homotopy Theory of Immersions

Given a differentiable immersion of a compact manifold M^n into $\boldsymbol{R}^{n+k}$, $f\colon M^n \to \boldsymbol{R}^{n+k}$, one can construct a k-dimensional normal vector bundle, ν_f, having the property that

$$\tau(M^n) \oplus \nu_f \cong \varepsilon_{n+k}.$$

Here $\tau(M^n)$ is the tangent bundle, "$\oplus$" denotes the Whitney sum operation, and ε_{n+k} denotes the trivial $(n+k)$-dimensional vector bundle. The immersion theorem of M. Hirsch [22] states that the converse to this situation is true as well. That is, if ν is any k-dimensional vector bundle over M^n with the property that $\tau(M^n) \oplus \nu = \varepsilon_{n+k}$, then there exists an immersion $f\colon M^n \to \boldsymbol{R}^{n+k}$ with normal bundle isomorphic to ν. Furthermore, isomorphism classes of k-dimensional vector bundles over M^n can be classified as the set of based homotopy classes of maps of M^n to the Grassmannian of k-planes in $\boldsymbol{R}^\infty$, $BO(k)$. Thus Hirsch's theorem can be viewed as a translation of the problem in Differential Topology of finding immersions of M^n into $\boldsymbol{R}^{n+k}$ into a problem of homotopy theory.

This homotopy-theoretic problem can be stated as follows: Let $BO = \varinjlim_k BO(k)$ and let $\nu_M\colon M^n \to BO$ classify the stable normal bundle of M^n. ν_M can be thought of as the classification map of the normal bundle of an embedding of M^n into a large-dimensional Euclidean space. Then Hirsch's theorem may be interpreted to state the following.

THEOREM [22]. *There exists an immersion $f\colon M^n \to \boldsymbol{R}^{n+k}$ if and only if there is a map $\nu\colon M^n \to BO(k)$ making the following diagram homotopy*

* During the preparation of this work the author was supported by NSF Grant MCS 82-03806 and by a fellowship from the A. P. Sloan Foundation.

commute:

$$\begin{tikzcd} & BO(k) \arrow[d] \\ M^n \arrow[ur, "\nu"] \arrow[r, "\nu_M"'] & BO. \end{tikzcd}$$

Thus Hirsch's theorem reduces the problem of the existence of immersions to a homotopy-theoretic lifting problem. This reduction has proved quite beneficial to both the field of Differential Topology and to that of Homotopy Theory. The study of these types of problems has widened the scope of homotopy theory and has led to the development of techniques that have found many applications to other areas of algebraic topology and geometry. Conversely, the infusion of homotopy-theoretic techniques into differential topology has resulted in the solution of many problems that would have been much more difficult to solve directly in a geometric or analytic manner.

One such problem is that of finding the smallest integer k_n with the property that every compact n-manifold immerses in $\boldsymbol{R}^{n+k_n}$. Using homotopy-theoretic techniques the following has recently been proved.

THEOREM [17]. *For* $n > 1$, $k_n = n - a(n)$, *where* $a(n)$ *is the number of ones in the dyadic expansion of* n.

The proof of this theorem in [17] was the final step in a twenty year long program devised by E. H. Brown Jr. and F. P. Peterson, which included contributions by many mathematicians. In this article I will describe some of the homotopy-theoretic constructions and techniques developed in this program, outline how they were used to prove the theorem, and also describe some of their applications to other problems in algebraic topology. In particular, in Section 2 I will discuss Brown–Gitler spectra and some of the many applications they have had to date.

In this paper, unless otherwise indicated, all (co)homology will be taken with $\boldsymbol{Z}_2$ coefficients, and by the term "n-manifold" I will mean a compact, n-dimensional, C^∞-manifold.

§ 1. History of the Immersion Conjecture and the program of its proof

In two celebrated papers appearing in 1944 [34, 35] H. Whitney proved, using geometric techniques, that every n-manifold embeds in $\boldsymbol{R}^{2n}$ and immerses in $\boldsymbol{R}^{2n-1}$.

During the next fifteen years there was little progress in constructing immersions of manifolds into Euclidean space, but it was observed that the first (primary) obstructions one encounters when trying to immerse an n-manifold M^n into $\boldsymbol{R}^{n+k}$ are the normal Stiefel–Whitney characteristic classes $\bar{w}_i(M^n)$, for $i > k$. Using the formulae of Wu [36] concerning how the normal and tangential characteristic classes of a manifold are related and how to compute them, Massey, in a paper appearing in 1960 proved the following:

THEOREM 1.1 [25]. $\bar{w}_i(M^n) = 0$ *for* $i > n - \alpha(n)$.

The fact that this theorem is best possible can be seen by the following well-known example.

If $n = 2^j$ then $\alpha(n) = 1$ and a simple calculation yields that $\bar{w}_{2^j-1}(\boldsymbol{R}P^{2^j}) \neq 0$. This in particular says that $\boldsymbol{R}P^{2^j}$ does not immerse in $\boldsymbol{R}^{2^{j+1}-2}$ and hence Whitney's immersion result is best possible in this case.

For the general case, write n in its dyadic expansion:

$$n = 2^{i_1} + 2^{i_2} + \ldots + 2^{i_k},$$

where $i_1 < i_2 < \ldots < i_k$, so that $\alpha(n) = k$. Let $M^n = \boldsymbol{R}P^{2^{i_1}} \times \ldots \times \boldsymbol{R}P^{2^{i_k}}$. Then using the Cartan formula for Stiefel–Whitney classes and the fact that $n - \alpha(n) = \sum_{j=1}^{k} (2^{i_j} - 1)$, one sees that $\bar{w}_{n-\alpha(n)}(M^n) \neq 0$.

After recalling that $H^*BO(k) = \boldsymbol{Z}_2[w_1, \ldots, w_k]$ and that $H^*BO = \boldsymbol{Z}_2[w_i\colon i > 0]$, Massey's result can be interpreted to say that if $\nu_M\colon M^n \to BO$ classifies the stable normal bundle of M^n, then there is a homomorphism $\bar{\nu}^*\colon H^*BO\big(n - \alpha(n)\big) \to H^*M^n$ making the following diagram commute:

$$\begin{array}{ccc} H^*BO & \xrightarrow{\nu_M^*} & H^*M^n \\ \downarrow & & \uparrow \bar{\nu}^* \\ H^*BO\big(n-\alpha(n)\big). & & \end{array}$$

What has become known as the *Immersion Conjecture* dating from Massey's theorem, is that the homomorphism $\bar{\nu}^*\colon H^*BO\big(n - \alpha(n)\big) \to H^*M^n$ can be realized by a map $\bar{\nu}\colon M^n \to BO\big(n - \alpha(n)\big)$ that homotopy lifts $\nu_M\colon M^n \to BO$. In view of Hirsch's immersion theorem this would imply that M^n immerses in $\boldsymbol{R}^{2n-\alpha(n)}$.

In 1963 E. Brown and F. Peterson completed the first step in a program that would eventually lead to a proof of this conjecture [5]. Roughly, their idea was to find a "universal" space together with a stable bundle that would encode the necessary bundle-theoretic properties of normal bundles of n-manifolds, thus alleviating the necessity of studying one manifold at a time.

One property this universal space would have is that its cohomology would be H^*BO, modulo the ideal I_n of all relations among the Stiefel–Whitney classes of normal bundles of n-manifolds. More precisely, let $\nu_M\colon M^n \to BO$ be a normal bundle map and let $I(M^n) \subset H^*BO = \mathbf{Z}_2[w_i\colon i > 0]$ be the kernel of the homomorphism $\nu_M^*\colon H^*BO \to H^*M^n$. Finally, define I_n to be the intersection of all such ideals,

$$I_n = \bigcap_{M^n} I(M^n).$$

In [5] Brown and Peterson computed I_n explicitly.

Observe that by the definition of I_n, every normal bundle homomorphism $\nu_M^*\colon H^*BO \to H^*M^n$ factors through the quotient module $\tilde{\nu}_M^*\colon H^*BO/I^n \to H^*M^n$, and that I_n is the largest ideal with this property. Moreover, in view of Massey's result we have that $w_i \in I_n$ for $i > n - a(n)$, so the projection $\varrho^*\colon H^*BO \to H^*BO/I_n$ factors through a homomorphism $\varrho_n^*\colon H^*BO(n-a(n)) \to H^*BO/I_n$. We may summarize this by saying that for any n-manifold M^n we have a commutative diagram

$$\begin{tikzcd}
H^*BO \arrow[r, "i^*"] \arrow[d, "\nu_M^*"'] \arrow[dr, "\varrho^*"] & H^*BO(n-a(n)) \arrow[d, "\varrho_n^*"] \\
H^*M^n & H^*BO/I_n. \arrow[l, "\tilde{\nu}_M^*"]
\end{tikzcd}$$

The two commutative triangles in this diagram yield two realization problems that form the basis of Brown and Petersons' program for a proof of the Immersion Conjecture. That is, their program may be broken into the following two steps.

(1.2). *Step 1*: Construct a space BO/I_n together with a map $\varrho\colon BO/I_n \to BO$ that has the following properties:

(a) $H^*(BO/I_n) = (H^*BO)/I_n$, and the homomorphism $\varrho^*\colon H^*BO \to H^*(BO/I_n)$ is the natural projection.

(b) For every n-manifold M^n there exists a map $\tilde{\nu}_M\colon M^n \to BO/I_n$

making the following diagram homotopy commute:

$$\begin{tikzcd} & BO/I_n \arrow[d, "\varrho"] \\ M^n \arrow[ur, "\tilde{\nu}_M"] \arrow[r, "\nu_M"'] & BO. \end{tikzcd}$$

Step 2: Show that there exists a map ϱ_n: $BO/I_n \to BO(n-a(n))$ making the following diagram homotopy commute:

$$\begin{tikzcd} & BO(n-a(n)) \arrow[d, "i"] \\ BO/I_n \arrow[ur, "\varrho_n"] \arrow[r, "\varrho"'] & BO. \end{tikzcd}$$

Step 1 was completed by Brown and Peterson in a paper appearing in 1979 [8]. Step 2 was recently completed by the author [16, 17].

Observe that these two results taken together with Hirsch's immersion theorem imply the Immersion Conjecture, since for any n-manifold M^n we can define an $(n-a(n))$-dimensional normal bundle map $M^n \to BO(n-a(n))$ to be the composition $M^n \xrightarrow[\tilde{\nu}_M]{} BO/I_n \xrightarrow[\varrho_n]{} BO(n-a(n))$.

As mentioned above, the first bit of progress in the Brown–Peterson program (1.2) was the calculation of the ideal $I_n \subset H^*BO$ [5]. Rather than describe I_n explicitly we shall describe its Thom isomorphic image $\Phi(I_n) \subset H^*(MO)$, where MO is the Thom spectrum of the universal stable vector bundle over BO. Actually we will describe the quotient group $H^*MO/\Phi(I_n)$.

To do this, we first recall the calculation of R. Thom [33] of the unoriented cobordism ring, η_*. Thom proved that

$$\eta_* \cong \pi_*MO \cong \mathbf{Z}_2[b_i\colon i \neq 2^j-1],$$

the graded polynomial algebra on generators b_i of grading i, so long as i is not of the form 2^j-1. Let $\omega = (i_1, \ldots, i_r)$ be a sequence of positive integers, none of which is one less than a power of two. Let $b_\omega = b_{i_1} \ldots b_{i_r}$ be the corresponding monomial in π_*MO and let $|\omega|$ denote the dimension, $|\omega| = i_1 + \ldots + i_r$. In [33] Thom actually proved the following.

THEOREM 1.3 [33]. *$MO \simeq \bigvee_{\omega} \Sigma^{|\omega|}KZ_2$, where "$\vee$" denotes the wedge sum of spectra, $\Sigma^{|\omega|}KZ_2$ is the $|\omega|$-fold suspension of the Z_2-Eilenberg–MacLane spectrum, so that*

$$\pi_q \Sigma^{|\omega|} KZ_2 = \begin{cases} Z_2 & \text{if } q = |\omega|, \\ 0 & \text{otherwise}. \end{cases}$$

The wedge is taken over all monomials $b_\omega \in \pi_* MO$.

Let A denote the mod2 Steenrod algebra. The following is then immediate.

COROLLARY 1.4. *$H^*MO = \bigoplus_{\omega} \Sigma^{|\omega|}A$, as A-modules.*

We can now state Brown and Peterson's Theorem [5].

THEOREM 1.5 [5]. *$H^*MO/\Phi(I_n) = \bigoplus_{\omega} \Sigma^{|\omega|}A/J_{[n-|\omega|/2]}$ where $J_k \subset A$ is the left ideal*

$$J_k = A\{\chi(\mathrm{Sq}^i)\colon i > k\}$$

and χ is the canonical antiautomorphism of A.

The next step in the Brown–Peterson program was to topologically realize these groups and homomorphisms by spaces (or spectra) and continuous maps. The first such realization was the construction of Brown–Gitler spectra whose cohomology realizes the A-modules A/J_k. These spectra and some of their many applications to homotopy theory will be discussed in the next section.

§ 2. Brown–Gitler spectra and their applications

In a paper appearing in 1973 [4] Brown and Gitler proved the following:

THEOREM 2.1. *For each integer $k \geqslant 0$ there is a spectrum $B(k)$ satisfying the following properties:*

(1) *$H^*B(k) \cong A/J_k$ as A-modules.*

(2) *If X is any space, the map $j_*\colon \pi_q(B(k)\wedge X) \to \pi_q(KZ_2 \wedge X) \cong H_q(X)$ is surjective for $q \leqslant 2k+1$. Here $j\colon B(k) \to KZ_2$, thought of as a cohomology class, is the generator of $H^*B(k)$ as an A-module.*

Furthermore, it was proved in [7] that these two properties completely characterize the 2-local homotopy type of $B(k)$.

We begin this section by discussing how these Brown–Gitler spectra were used in the proof of the Immersion Conjecture, and will later describe some of the many other problems in algebraic topology to which they've recently been applied.

Let M^n be an n-manifold, and let $T\nu(M^n)$ be the Thom spectrum of the stable normal bundle of M^n. Now if E is any spectrum, let $E_*(X)$ and $E^*(X)$ denote the generalized homology and cohomology of X with coefficients in E. Then by the duality between M^n and $T\nu(M^n)$ we have isomorphisms $E^q(T\nu(M^n)) \cong E_{n-q}(M^n_+)$. Applying this duality to part (2) of Theorem 2.1 we have the following.

COROLLARY 2.2. *$B(k)^q(T\nu(M^n)) \to H^q(T\nu(M^n))$ is surjective for $n-q \leqslant 2k+1$. Equivalently, if α: $T\nu(M^n) \to \Sigma^q KZ_2$ represents any cohomology class for $n-q \leqslant 2k+1$, then there is a map $\tilde{\alpha}$: $T\nu(M^n) \to \Sigma^q B(k)$ so that the composition $T\nu(M^n) \xrightarrow[\tilde{\alpha}]{} \Sigma^q B(k) \xrightarrow[j]{} \Sigma^q KZ_2$ is homotopic to α.*

We observe that this corollary allows the completion of the Thom spectrum version of step (1) in the Brown–Peterson program (1.2) for the proof of the Immersion Conjecture. More precisely, we may define the spectrum

$$MO/I_n = \bigvee_{|\omega| \leqslant n} \Sigma^{|\omega|} B[n-|\omega|/2]$$

and a map of spectra τ: $MO/I_n \to MO \simeq \bigvee_{\omega} \Sigma^{|\omega|} KZ_2$ to be the wedge of the suspensions of the generating maps j: $B(k) \to KZ_2$.

COROLLARY 2.3. (1) *$H^*(MO/I_n) \cong H^*MO/\Phi(I_n)$.*

(2) *If t_M: $T\nu(M^n) \to MO$ is the map of Thom spectra induced by the stable normal bundle map ν_M, then there is a map of spectra $\tilde{t}_M$: $T\nu(M^n) \to MO/I_n$ making the following diagram homotopy commute:*

$$\begin{tikzcd} & MO/I_n \arrow[d, "\tau"] \\ T\nu(M^n) \arrow[ur, "\tilde{t}_M"] \arrow[r, "t_M"'] & MO. \end{tikzcd}$$

Proof. Part (1) follows from Theorems 2.1 and 1.5. Part (2) follows from Corollary 2.2 since MO is a wedge of Eilenberg–MacLane spectra (1.3).

Now in a paper appearing in 1971 [10] R. Brown proved that every n-manifold is cobordant to one that immerses in $\boldsymbol{R}^{2n-\alpha(n)}$. Using the translation of this theorem to a statement about the homotopy groups of certain Thom spectra given by the Thom–Pontryagin theorem, Brown and Peterson were able to use this to prove the following [6]:

THEOREM 2.4. *There exists a map* $\tau_n\colon MO/I_n \to MO(n-\alpha(n))$ *that homotopy lifts* $\tau\colon MO/I_n \to MO$.

Notice that this completes the Thom spectrum version of step (2) of the Brown–Peterson program (1.2).

The object now is therefore to "de-Thom-ify" Theorems 2.3 and 2.4; that is, to complete program (1.2). The first indication that this might be possible were theorems of Brown and Peterson [7] and the author [14] stating that Brown–Gitler spectra can be realized as the Thom spectra of certain vector bundles. The example of [7] can be described as the vector bundle V_k defined as

$$V_k\colon F(\boldsymbol{R}^2, k) \times_{\Sigma_k} \boldsymbol{R}^k \to F(\boldsymbol{R}^2, k)/\Sigma_k$$

where $F(\boldsymbol{R}^2, k) = \{(x_1, \ldots, x_k)\colon x_i \in \boldsymbol{R}^2, \text{ and } x_i \neq x_j \text{ if } i \neq j\}$, and Σ_k is the symmetric group on k letters. It is proved in [7] that the Thom spectrum TV_k is 2-locally homotopy equivalent to $B[k/2]$.

We remark that in both [7] and [14] the representations of Brown–Gitler spectra as Thom spectra were inspired by calculations of their cohomology by Mahowald in [24].

Although the results of [7] and [14] describe Brown–Gitler spectra as Thom spectra, it was not clear (and is still not clear) how to patch these bundles together to yield a space BO/I_n and stable vector bundle ϱ whose Thom spectrum is MO/I_n. Thus a more obstruction-theoretic approach was taken and will be outlined in Section 3. We end this section by describing how Brown–Gitler spectra have been applied to other problems in algebraic topology.

Applications. (1) A standard observation yields that the configuration space $F(\boldsymbol{R}^2, k)/\Sigma_k$ is an Eilenberg–MacLane space of type $K(\beta_k, 1)$ where β_k is Artin's braid group on k-strings (see [1] for instance). It is easy to see that the permutation representation of β_k given by sending a braid to the associated permutation of the endpoints of the strings classifies the bundle V_k over $K(\beta_k, 1)$. Thus using the equivalence $TV_k \simeq B[k/2]$, the calculation of the homotopy groups of $B[k/2]$ in [7, 14] classifies, up to cobordism, braid group oriented manifolds. This point of view has been exploited by Sanderson [28] and Bullett [11].

(2) The author's work in [14] was generalized in [15] to give an axiomatic description of Brown–Gitler spectra in terms of the product structure they possess.

(3) Brown–Gitler spectra, and particularly the representations as the Thom spectra TV_k, were used by Mahowald [24] to produce an infinite

family in the homotopy groups of spheres, $\pi_*(S^0)$. (See also [9, 19] for simplifications making stronger use of the homotopy-theoretic properties of Brown–Gitler spectra.)

(4) In [13] the author constructed odd primary analogues of Brown–Gitler spectra. Using these and techniques of Mahowald he produced infinite families in each p-primary component of $\pi_*(S^0)$ (p an odd prime). Then by studying exponents of certain Brown–Gitler spectra the author and Goerss completed the classification of mod p ($p \geqslant 5$) secondary cohomology operations that can act nontrivially on complexes of the form $S^n \cup_f D^m$ [18].

(5) In [19] the author, Jones, and Mahowald generalized work of Browder [2] and Brown [3] in order to give necessary and sufficient homotopy-theoretic conditions for a manifold immersed in a low-codimensional Euclidean space to have nonzero Kervaire invariant. Using Brown–Gitler spectra theory they were able to show that for every $j > 0$ there is an oriented manifold of dimension $2^{j+1}-2$ immersed in $\boldsymbol{R}^{2^{j+1}}$ with nonzero Kervaire invariant.

(6) By studying the Spanier–Whitehead duals of Brown–Gitler spectra, Carlsson was able to prove the Segal–Burnside ring conjecture for elementary abelian 2-groups [12]. H. Miller then used these techniques and an impressive unstable Adams spectral sequence argument to prove a famous conjecture of Sullivan, stating that if G is a finite group and X a finite CW-complex, then any map $f\colon BG \to X$ is null homotopic [26, 27].

(7) The ideal $J_k = A\{\chi(\mathrm{Sq}^i)\colon i > k\}$ is the image under χ of all operations $a \in A$ having the property that if $x \in H^q(X)$ is any cohomology class of a space X such that $2(\dim a + q) > k$, then $a(x) = 0$. This ideal has obvious analogues in the algebras of operations of other (generalized) cohomology theories. Thus one can ask for analogues of Brown–Gitler spectra with respect to these theories. These have been developed for mod p (p odd) cohomology, integral cohomology, and real connected K-theory in [13, 29, 20, 21]. In view of the many applications the original Brown–Gitler spectra have had, it is reasonable to expect that these generalized versions (and others yet to be constructed) will have a similar impact on homotopy theory.

§ 3. De-Thom-ifications

In this section we sketch how Theorems 2.3 and 2.4 were de-Thom-ified in [8] and [17]. That is, we describe how program (1.2) for the proof of the Immersion Conjecture was completed.

By the results of [4] a good deal of information was known about the Adams spectral sequence for the Brown–Gitler spectrum $B(k)$. Taking the wedge of the Adams resolutions of the appropriate Brown–Gitler spectra, Brown and Peterson obtained a tower, in effect an Adams resolution, for MO/I_n:

$$\begin{array}{ccccccccc} MO/I_n \to \dots \to & T_{i+1} & \to & T_i & \to \dots \to & T_0 & = MO \\ & & & \downarrow & & \downarrow & \\ & & & L_i & & L_0 & \end{array} \tag{3.1}$$

where each $T_{i+1} \to T_i \to L_i$ is a stable fibration sequence with L_i a wedge of Eilenberg–MacLane spectra. In [8] Brown and Peterson succeeded in de-Thom-ifying this tower. That is, they showed that there is a tower

$$\begin{array}{ccccccc} \to \dots \to B_{i+1} \to & B_i & \to \dots \to & B_0 & = BO \\ & \downarrow & & \downarrow & \\ & K_i & & K_0 & \end{array} \tag{3.2}$$

of fibrations of spaces having the property that there is an n-dimensional homotopy equivalence of spectra $TB_i \simeq T_i$, where TB_i is the Thom spectrum. They proved this by induction on i. The inductive step was completed with the help of calculations of Mahowald [23] and Browder [2] concerning how the cohomology of a Thom spectrum changes when a cohomology class in the base space is killed. These calculations were used to identify the precise obstruction to T_{i+1} being a Thom spectrum under the assumption that T_i is. They then succeeded in computing these obstructions and showing that they are zero.

BO/I_n was defined to be, more or less, the inverse limit of the B_i's. By comparing tower (3.2) to tower (3.1) via the Thom isomorphism, and using the Adams spectral sequence results of [4], they were then able to prove the following, which completes step 1 of program (1.2).

THEOREM 3.3 [8]. *For any n manifold M^n, the stable normal bundle map $\nu_M\colon M^n \to BO$ homotopy lifts up tower* (3.2) *to yield a map $\tilde{\nu}\colon M^n \to BO/I_n$.*

The proof of the Immersion Conjecture was completed by the author in [17] when he proved the following:

THEOREM 3.4. *The map $\varrho\colon BO/I_n \to BO$ homotopy lifts to a map $\varrho_n\colon BO/I_n \to BO(n-a(n))$.*

The de-Thom-ification obstruction theory described above was used very heavily in the proof of Theorem 3.4. Indeed, this theory, together with a theorem of Snaith stating that each $BO(2k)$ is a stable homotopy retract of BO [31], were the two main components in the first step of the proof of Theorem 3.4 (Chapter 1 of [17]), which was to reduce this theorem to the following lemma.

LEMMA 3.5. *Let P_n be the homotopy pull-back of the maps $i\colon BO(n-a(n))\to BO$ and $\varrho\colon BO/I_n\to BO$. Then there is a space X_n together with a map $h_n\colon X_n\to P_n$ satisfying the following properties:*

(1) *If f_n and g_n are the compositions $X_n\xrightarrow{h_n}P_n\to BO(n-a(n))$ and $X_n\xrightarrow{h_n}P_n\to BO/I_n$ respectively, then there is a splitting map of 2-local Thom spectra, $\sigma_n\colon MO/I_n\to TX_n$. That is, $1\simeq Tg_n\circ\sigma_n\colon MO/I_n\to TX_n\to MO/I_n$.*

(2) *The following diagram of 2-local Thom spectra homotopy commutes:*

$$\begin{array}{ccc} TX_n & \xrightarrow{Th_n} & TP_n \\ {\scriptstyle Tg_n}\downarrow & & \uparrow{\scriptstyle Th_n} \\ MO/I_n & \xrightarrow[\sigma_n]{} & TX_n. \end{array}$$

The point of reducing 3.4 to 3.5 was to make use of the considerably more information about the Thom spectrum MO/I_n than was available about the base space BO/I_n. In particular, X_n was a space of the form

$$X_n=\coprod_{\omega} M_\omega\times\big(F(\mathbf{R}^2,n-|\omega|)/\Sigma_{n-|\omega|}\big),$$

where M_ω is a manifold representing the cobordism class $b_\omega\in\pi_*MO$. Moreover, the stable vector bundle over X_n was taken to be the disjoint union of the product of the stable normal bundle over M_ω and the bundle $V_{n-|\omega|}$ over $F(\mathbf{R}^2,n-|\omega|)/\Sigma_{n-|\omega|}$.

Lemma 3.5 was proved by induction on n. A major step was to construct pairings $BO/I_r\times BO/I_s\to BO/I_{r+s}$ compatible with the Whitney sum pairings of vector bundles, and so that on the Thom spectrum level these pairings are those induced by the pairings of Brown–Gitler spectra coming from the natural pairings of braid groups (see Chapter 4 of [17]). As one might imagine, de-Thom-ification obstruction theory was heavily used for this.

In the case when n is of the form $n=2^r-1$, so that there are no indecomposables of dimension n in the cobordism ring, the above pairings quickly lead to a completion of the inductive step (Chapter 2 of [17]). When $n\neq 2^r-1$, the n-dimensional indecomposable cobordism class had to be accounted for using a different argument (Chapter 3 of [17]) again making heavy use of de-Thom-ification obstruction theory.

It seems clear that this obstruction theory, Brown–Gitler spectra, and the other constructions and techniques mentioned above, will continue to prove quite useful in other areas of algebraic topology.

References

[1] Birman J., *Braids, Links, and Mapping Class Groups*, Annals of Math. Studies 82, Princeton Univ. Press, 1974.

[2] Browder W., The Kervaire Invariant of Framed Manifolds and its Generalizations, *Ann. of Math.* **90** (1969), pp. 157–186.

[3] Brown Jr. E. H., Generalizations of the Kervaire Invariant, *Ann. of Math.* **95** (1972), pp. 368–383.

[4] Brown Jr. E. H. and Gitler S., A Spectrum whose Cohomology is a Certain Cyclic Module over the Steenrod Algebra, *Topology* **12** (1973), pp. 283–295.

[5] Brown Jr. E. H. and Peterson F. P., Relations among Characteristic Classes I, *Topology* **3**, Suppl. 1 (1964), pp. 39–52.

[6] Brown Jr. E. H. and Peterson F. P., On Immersions of n-Manifolds, *Advances in Math.* **24** (1977), pp. 74–77.

[7] Brown Jr. E. H. and Peterson F. P., On the Stable Decomposition of Ω^2S^{2+r}, *Trans. Amer. Math. Soc.* **243** (1978), pp. 287–298.

[8] Brown Jr. E. H. and Peterson F. P., A Universal Space for Normal Bundles of n-Manifolds, *Comment. Math. Helv.* **54** (1979), pp. 405–430.

[9] Brown Jr. E. H. and Peterson F. P., *The Brown–Gitler spectrum, Ω^2S^3 and $\eta_j \in \pi_{2^j}(S)$*, preprint, 1979.

[10] Brown R. L., Immersions and Embeddings up to Cobordism, *Canad. J. Math.* **23** (6) (1971), pp. 1102–1115.

[11] Bullett S., Braid Orientations and Stiefel–Whitney Classes, *Quart. J. Math. Oxford* (2) **32** (1981), pp. 267–285.

[12] Carlsson G., G. B. Segal's Burnside Ring Conjecture for $(\mathbf{Z}/2)^k$, *Topology* **22** (1983), pp. 83–103.

[13] Cohen R. L., *Odd Primary Infinite Families in Stable Homotopy Theory*, Memoirs of A.M.S. 242 (1981).

[14] Cohen R. L., The Geometry of Ω^2S^3 and Braid Orientations, *Invent. Math.* **54** (1979), pp. 53–67.

[15] Cohen R. L., Representations of Brown–Gitler Spectra. In: Proc. Top. Symp. at Siegen, 1979, *Springer Lecture Notes* 788 (1980), pp. 399–417.

[16] Cohen R. L., Immersions of Manifolds, *Proc. Nat. Acad. Sci. USA* **79** (1982), pp. 3390–3392.

[17] Cohen R. L., *The Immersion Conjecture for Differentiable Manifolds*, to appear.

[18] Cohen R. L. and Goerss P., *Secondary Cohomology Operations that Detect Homotopy Classes*, to appear.

[19] Cohen R. L., Jones J., and Mahowald M., *The Kervaire Invariant of Immersions*, to appear.

[20] Goerss P., Ph.D. dissertation, M.I.T., 1983.

[21] Jones J. and Mahowald M., Private communication, 1983.
[22] Hirsch M. W., Immersions of Manifolds, *Trans. Amer. Math. Soc.* **93** (1959), pp. 242–276.
[23] Mahowald M., On Obstruction Theory in Orientable Fibre Bundles, *Trans. Amer. Math. Soc.* **110** (1964), pp. 315–349.
[24] Mahowald M., A New Infinite Family in ${}_2\pi_*^s$, *Topology* **16** (1977), pp. 249–256.
[25] Massey W. S., On the Stiefel-Whitney Classes of a Manifold, *Amer. J. Math.* **82** (1960), pp. 92–102.
[26] Miller H., *The Sullivan Fixed-Point Conjecture and Brown–Gitler Spectra*, to appear.
[27] Miller H., *The Sullivan Conjecture*, to appear.
[28] Sanderson B., The Geometry of Mahowald Orientations. In: Proc. Aarhus Conf. in Alg. Top., 1978, *Springer Lecture Notes* 763, pp. 152–174.
[29] Shimamoto D., Ph.D. dissertation, Brandeis University, 1982.
[30] Snaith V., A Stable Decomposition for $\Omega^n S^n X$, *J. London Math. Soc.* **2** (1974), pp. 577–583.
[31] Snaith V., *Algebraic Cobordism and K-Theory*, Memoirs of A. M. S. 221 (1979).
[32] Spanier E., Function Spaces and Duality, *Ann. of Math.* **70** (1959), pp. 338–378.
[33] Thom R., Quelques propriétés globales des variétés différentiables, *Comment. Math. Helv.* **28** (1954), pp. 17–86.
[34] Whitney H., The Self-Intersections of a Smooth n-Manifold in $2n$-Space, *Ann. of Math.* **45** (1944), pp. 220–246.
[35] Whitney H., The Singularities of a Smooth n-Manifold in $(2n-1)$-Space, *Ann. of Math.* **45** (1944), pp. 247–293.
[36] Wu W. T., Classes caractéristiques et i-carrés d'une variété, *C. R. Acad. Sci. Paris* **230** (1950), pp. 508–511.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

S. K. DONALDSON

Gauge Theory and Topology

We shall describe in this talk a recent application of some of the mathematical techniques developed for the study of non-abelian field theories to a problem in 4-manifold topology. Details will be found in [2] with a more leisurely account in [4]. Two related talks at this congress are those by Freedman and Uhlenbeck.

1. Intersection forms

For any oriented manifold M^{4m} there is a symmetric, bilinear intersection form

$$H^{2m}(M)/\text{Torsion}\times H^{2m}(M)/\text{Torsion}\to \boldsymbol{Z}.$$

For each integer $m>1$ and each symmetric form of determinant ± 1 there is a $2m-1$ connected PL manifold of dimension $4m$ realizing that form. The corresponding statement for $m=1$ is false; any PL 4-manifold has a smooth structure and then the theorem of Rohlin [5] gives a restriction. The further restriction we report on here is unusual in that it appears not to have any analogue for higher dimensional manifolds:

THEOREM. *The only positive definite forms realized by smooth simply connected 4-manifolds are the standard ones, given by the identity matrix in some base.*

Since spin manifolds have even forms, a trivial corollary to the theorem is: *if a smooth simply connected spin 4-manifold has a definite form then* $H_2(M)=0$. For brevity we shall describe just the proof of this corollary. In this case it is necessary and sufficient to show that any such manifold M has intersection form of signature zero which is the same as saying that there is a 5-manifold $\mathfrak{M}$ which is orientable and has boundary M (by the well-known classification of 4-manifolds up to cobordism). We shall find such a manifold $\mathfrak{M}$ 'in nature' as *the moduli space of self-dual* *connections on the $k = 1$ SU(2) bundle over M* once we have chosen some Riemannian metric on M.

2. Gauge theories on Riemannian 4-manifolds

The mathematical framework here is now well known. The electromagnetic potential is a 1-form A and the associated field $F = dA$ satisfies Maxwell's equations $d^* F = 0$. Transferred to a compact Riemannian manifold these ideas become part of the Hodge theory of harmonic forms. Now fix a compact Lie group G and generalize A to a 1-form with values in the Lie algebra $\mathfrak{g}$ of G. The field is then given by the non-linear expression:

$$F = dA + \tfrac{1}{2}[A, A].$$

Geometrically, A represents a connection on a principle G-bundle and F is its curvature. The generalization of Maxwell's equations are the *Yang–Mills equations* which are the same to highest order but have non-linear lower order terms coming from the bracket in $\mathfrak{g}$. Taking G to be the abelian group S^1, these terms vanish and we recover Maxwell's equations. Again all of this extends to bundles over Riemannian manifolds although not much is known about the solutions in general.

Special to four dimensions is the *conformal invariance* of the Yang–Mills equations and the existence of first order conditions which force the solutions of the second order Yang–Mills equations. These are the *self-duality equations* which require that the curvature be fixed by the * operator. These *self-dual connections* are easier to understand: on any compact Riemannian 4-manifold X the solutions, for the group SU(2) say, are indexed by an integer $k \geqslant 0$ corresponding to the topological type of the bundle P and for each value of k are parametrized by a finite-dimensional moduli space $\mathfrak{M}_k(X)$.

The prototype and explicit examples are the $k = 1$ *instantons* on S^4 or, using the conformal invariance $\boldsymbol{R}^4$. On $\boldsymbol{R}^4$ these each have an associated length scale and SO(4) symmetry about a centre in $\boldsymbol{R}^4$. As the scale tends to zero the field densities converge to a delta function at the centre. In this case the centre and scale completely determine the solution, so one can see the moduli space directly.

3. Analytical properties of self-dual connections

It is a familiar fact that the solutions to an elliptic linear differential equation over a compact manifold form a finite-dimensional vector space — that is, the unit ball in the space is compact. Similarly, if a solution is defined on the complement of a point and is p-integrable for large enough p then it extends smoothly over the whole manifold. In two papers [8], [9] Karen Uhlenbeck proved analogous theorems for the non-linear Yang–Mills equations, using the norm of the curvature to control a connection. The following corollary to these theorems was probably well known to workers in the area:

PROPOSITION. *For any sequence A_i of self-dual connections on a bundle P over X there is a finite set of points $\{x_1, x_2, \ldots, x_l\} \subset X$, a bundle P' with topological invariant k', $0 \leqslant k' \leqslant k-1$ and a self-dual connection A_∞ on P', such that there are bundle isomorphisms:*

$$\varrho_i\colon P'|_{X\setminus\{x_1,\ldots,x_l\}} \to P|_{X\setminus\{x_1,\ldots,x_l\}}$$

with some subsequence of the pulled back connections $\varrho_i^*(A_i)$ converging to A_∞, uniformly with all derivatives on compact subsets of $X\setminus\{x_1, x_2, \ldots, x_l\}$.

If $l > 0$ the curvature of the connections A_i gathers in integer units over the points x_s and the corresponding twists in the bundle are lost in the limit. In particular, if $k = l$ there is at most one point x_s and if there is one the bundle P' is the trivial bundle. If further the base manifold X is simply connected the limit A_∞ is the trivial product connection.

Conversely, Clifford Taubes showed that under very general hypotheses one could construct a self-dual connection over a 4-manifold X whose structure resembles that of an instanton near to a point $x \in X$ and which is close to the flat connection elsewhere. For the group SU(2) he proved:

THEOREM [6]. *If the compact Riemannian 4-manifold X carries no non-zero anti self-dual harmonic 2-forms, then there exist irreducible self-dual connections over X for each $k > 0$.*

Of course, this condition on the manifold is precisely that the intersection form be positive definite. More recently Taubes extended his methods to arbitrary 4-manifolds: self-dual connections exist if k is large compared to the dimension b_2^- of the negative part of the intersection form ([7]). Except for some special cases sharp non-existence results seem to be lacking at the moment.

4. Moduli spaces

The local analysis of moduli spaces of self-dual connections was initiated by Atiyah, Hitchin and Singer [1] and follows closely the Kodaira–Spencer–Kuranishi theory of deformation of complex structures. If A is a self-dual connection a neighbourhood of the corresponding point in the moduli space is a manifold so long as:

(a) The connection A is irreducible.

(b) A certain differential operator associated to A is surjective.

For our purposes (b) is unimportant. The usual transversality arguments apply in this context and one can deform away any such singularities (recently Uhlenbeck proved [4] that for generic Riemannian metrics these singularities are absent). Similarly for the $k = 1$ bundle over our spin manifold M there are no possible reductions for topological reasons so in this case we can suppose that the moduli space is smooth; the general theorem is proved by keeping track of singularities of type (a).

The dimension of this moduli space, subject to the qualifications above, is given by an application of the Atiyah–Singer Index theorem:

$$\dim \mathfrak{M}_k(x) = 8k - 3(1 - b_1 + b_2^-).$$

Note in passing that this formula helps one to understand the obstruction encountered by Taubes — if b_2^- is large compared with k and the first Betti number b_1 then $\dim \mathfrak{M}_k(x)$ is negative which means that solutions do not generically exist. Anyway for our manifold M solutions with $k = 1$ exist and are parametrized by a 5-dimensional manifold $\mathfrak{M} = \mathfrak{M}_1(M)$, also a further application of index theory shows that this manifold is orientable.

THEOREM. *Under our hypotheses the moduli space $\mathfrak{M} = \mathfrak{M}_1(M)$ contains as an open subset a collar $\cong M \times (0, 1)$. The complement of this collar in $\mathfrak{M}$ is compact, so $\mathfrak{M}$ can be compactified to give a null cobordism of M.*

Establishing this makes up the main technical burden of the proof, although the result is reasonably intuitive. The collar is formed by the 5-parameter family of Taubes' construction and one has to check that the centre and scale of these instanton like solutions give coordinates. The proof that the complement is compact follows the line of argument of Section 3 above.

References

[1] Atiyah M. F., Hitchin N. J., and Singer I. M., Self duality in four dimensional Riemannian geometry, *Proc. Roy. Soc. Lond.* **A362** (1978).

[2] Donaldson S. K., An application of gauge theory to four dimensional topology, *Journal of Diff. Geom.* **18** (1983).

[3] Freedman M. H., The topology of four dimensional manifolds, *Journal of Diff. Geom.* **17** (1982).

[4] Freed D., Freedman M. H., and Uhlenbeck K. K., *Gauge theories and 4-manifolds*, M.S.R.I. Berkeley preprint.
[5] Rohlin V. A., New results in the theory of four dimensional manifolds, *Dokl. Akad. Nauk. SSSR* **84** (1952).
[6] Taubes C. H., Self dual Yang Mills connections on non self dual 4-manifolds, *Journal of Diff. Geom.* **17** (1982).
[7] Taubes C. H., *Self dual connections on manifolds with indefinite intersection form*, to appear.
[8] Uhlenbeck K. K., Connections with L bounds on curvature, *Commun. Math. Phys.* **83** (1982).
[9] Uhlenbeck K. K., Removable singularities in Yang–Mills fields, *Commun. Math. Phys.* **83** (1968).

UNIVERSITY OF OXFORD,
MATHEMATICAL INSITUTE

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

MICHAEL H. FREEDMAN

The Disk Theorem for Four-Dimensional Manifolds

The two-dimensional disk D^2 seems to serve as a fundamental unit in manifold topology, mediating algebra and geometry. For manifolds of dimension greater than or equal to five, intersection pairings taking values in group rings $\boldsymbol{Z}[\pi_1 M]$ are crucial to the classification problem. The pairings are translated into precise geometric information by isotopies guided by imbedded two-disks. This is the "Whitney trick" [18], key to both s-cobordism and (even-dimensional) surgery theorems. The topology of three-dimensional manifolds is closely tied to the fundamental group by the classical disk locating theorems, Dehn's Lemma and the Loop Theorem. These theorems make the hierarchy theory run and eventually lead to toroidal decomposition. (And conversely, the least understood 3-manifolds are those having no fundamental group to decompose by imbedded disks — homotopy 3-spheres (?).) One could extend this pattern to dimension two by quoting the continuous-boundary-value Riemann mapping theorem (together with the uniformization theorem) as the 2-dimensional disk theorem.

There is now a 4-dimensional 2-disk imbedding theorem. Its simply connected version was the key to the work on the Poincaré conjecture [7]. The body of this paper is a discussion of its proof, with applications being given at the end. It has two restrictions which deserve comment before stating the theorem. First, it applies to the topological category. The analogous statement in the smooth category is false. This failure is a consequence of S. Donaldson's recent theorem [4] which states that the only definite form which can occur as the intersection form of a closed, smooth, simply connected 4-manifold is the standard form: $\pm$ identity matrix. On the other hand, the disk theorem readily implies a $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$-destabilization lemma, allowing the construction of necessarily topological manifolds with "exotic" definite forms. For example, one may use the identity of quadratic forms:

$$\bigoplus_{9} 1 \oplus \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cong E_8 \oplus 1 \oplus \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

to build a topological manifold with form $\cong E_8 \oplus 1$.

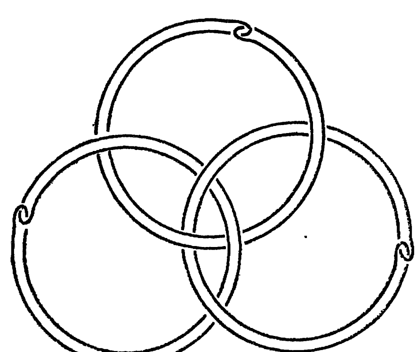
Naively, the disk theorem would mix categories; a smooth set up would yield a topological conclusion (and the reader should, initially, think of the theorem in this way). However, a long beautiful route leads to a fully topological statement. Frank Quinn's local version [15] of the (simply connected) mixed category disk theorem extends the smoothing theory of Kirby–Edwards, Kirby–Siebenmann, and Lashof–Rothenberg [10] sufficiently in dimension four to translate the topological hypothesis of the disk theorem into a smooth set-up.

The second restriction involves the fundamental group of the 4-manifold into which we are imbedding the disk. Here the situation is less tidy. It is not presently known if any fundamental group restriction is actually required for the disk theorem and its corrollaries. Having undergone several changes of heart, let me propose, with cautious neutrality:

CONJECTURE. *There is no topological imbedding*

$$i\colon \coprod_{3} (D^2 \times D^2, \partial D^2 \times D^2) \hookrightarrow (B^4, \partial B^4)$$

with $i(\coprod_{3} \partial D^3 \times 0) \subset S^3$ *being a 3-component link which is an (untwisted) Whitehead double of the Borromean Rings,* Wh(Bor).



Wh(Bor)

The conjecture that the above link is not "topologically slice" is at odds with surgery[1] and would require a failure of the disk theorem for free fundamental groups. This example lies at the heart of the fundamental group issue. I gratefully exploit this wide forum to advertise this open question.

DISK THEOREM 1. *Let $j\colon (D^2\times D^2, \partial D^2\times D^2)\to(M^4, \partial M^4)$ be a topological immersion of a 2-handle[2] which is an imbedding near $\partial D^2\times D^2$. Suppose that there exists another immersion of a 2-sphere $\times$ disk, $\alpha\colon S^2\times D^2\to M$ and that the intersections are transverse[3] and satisfy: $\mu(j) = 0$, $\mu(\alpha) = 0$, $\lambda(j, \alpha) = 1$. Assume $\pi_1(M)$ is good. Then j is topologically regularly homotopic (relative to a neighborhood of $\partial D^2\times D^2$) to a topological imbedding $i\colon (D^2\times D^2, \partial D^2\times D^2)\hookrightarrow(M^4, \partial M^4)$.*

Good fundamental groups are simply those for which the theorem can be proved. These include (in an order which follows the logic of the proof): $\pi_1\cong$ trivial group, $\pi_1\cong$ finite group, $\pi_1\cong$ integers, $\pi_1\cong$ Abelian, $\pi_1\cong$ solvable, and $\pi_1=$ infra-solvable (containing a solvable subgroup of finite index). The class of groups for which theorem has been proved may be described as: (1) containing all finite groups and the integers, and (2) being closed under: subgroup, quotient group, extension, and (infinite) nested union.

It appears to be more than accidental that groups of polynomial growth, which by a theorem of Gromov coincide with finitely generated infra-nilpotent groups, are included in this list.

[1] There is a surgery problem with a well-defined and vanishing obstruction in $L_4(\boldsymbol{Z}*\boldsymbol{Z}*\boldsymbol{Z})$ whose "solution" would construct a possible closed slice complement $\mathrm{cl}[B^4 - i(\amalg_3 D^2\times\tfrac{1}{2}D^2)]$. The disk theorem for $\pi_1(M)\cong\boldsymbol{Z}*\boldsymbol{Z}*\boldsymbol{Z}$ would permit the construction of such a solution.

[2] In the topological category a useful disk imbedding theorem should produce a 2-handle rather than a naked 2-disk. The normal coordinates are needed to define the tapering of the Whitney trick. Wild possibilities, analogous to the Alexander horned sphere in 3-space, are to be avoided.

[3] Quinn's version of smoothing [15] allows topological surfaces to be put in the usual normal crossing form by a small regular homotopy. This done, the Wall forms hypothesis, above, may be stated as:

(1) the double points of j may be arranged in pairs (p_i, p_i') so that a "Whitney circle", C_i, [= a circle on $j(D^2\times 0)$ running through p_i and p_i' once each and disjoint from p_j and p_j' for $j\neq i$] is null homotopic in M and p_i and p_i' acquire opposite signs from the local orientation induced by the null homotopy;

(2) the double points of $\alpha(S^2\times 0)$ are similarly paired; and

(3) all intersection points of $j(D^2\times 0)$ and $\alpha(S^2\times 0)$ but one are arranged in such pairs.

The disk theorem has as corrollaries appropriate s-cobordism and surgery theorems. In both cases, the dual class α in the hypothesis arises naturally in the course of the proof. For simplicity, noncompact and relative statements will be avoided, although such versions exist.

COROLLARY 1. *Let* $(W^5;\ \partial_+ W,\ \partial_- W)$ *be a compact topological s-cobordism. If* $\pi_1(W^5)$ *is good then* W^5 *is homeomorphic to* $\partial_+ W \times [0, 1]$.

Note. Quinn's paper [15] contains a remarkable proof of the 4-dimensional annulus conjecture based on a local version of the disk theorem for $\pi_1 = 0$. Using the disk theorem for $\pi_1 = \boldsymbol{Z}^4$, Kirby's original proof of the annulus conjecture can be extended to dimension four by using the noncompact extension of the above theorem to recognize P. L. structures on the punctured four-torus, T^4—pt. (See Section 10 of [7] for details of the noncompact case.)

COROLLARY 2. *Let* $f\colon (\overset{\substack{\nu\\ |}}{M},\partial)\to(\overset{\substack{\zeta\\ |}}{X},\partial)$ *be a degree one normal map from a 4-manifold with boundary to a Poincaré pair. (Normal means that* ζ *is a topological* $\boldsymbol{R}^N$ *bundle over* X *which pulls back to the stable normal bundle of* M.) *If* $f|_{\partial M}\colon \partial M\to\partial X$ *is a (simple) homology isomorphism with* $\boldsymbol{Z}[\pi_1 X]$ *coefficients then there is defined an obstruction* $\theta\in L_4^{h(s)}(\pi_1 X)$ *to constructing a normal bordism* (rel ∂) $\bar f$ *from* f *to* f' *where* $f'\colon (M', \partial)\to(X, \partial)$ *is a (simple) homotopy equivalence. If the obstruction vanishes* $(\theta = 0)$ *and* $\pi_1(X)$ *is good then* $(\bar f; f, f')$ *can be constructed.*

There are nontrivial special cases where surgery can be completed in spite of a bad (= not good) fundamental group. Historically interesting, in view of the role of the Whitehead continuum in the solution to the 4-dimensional Poincaré conjecture [7] is

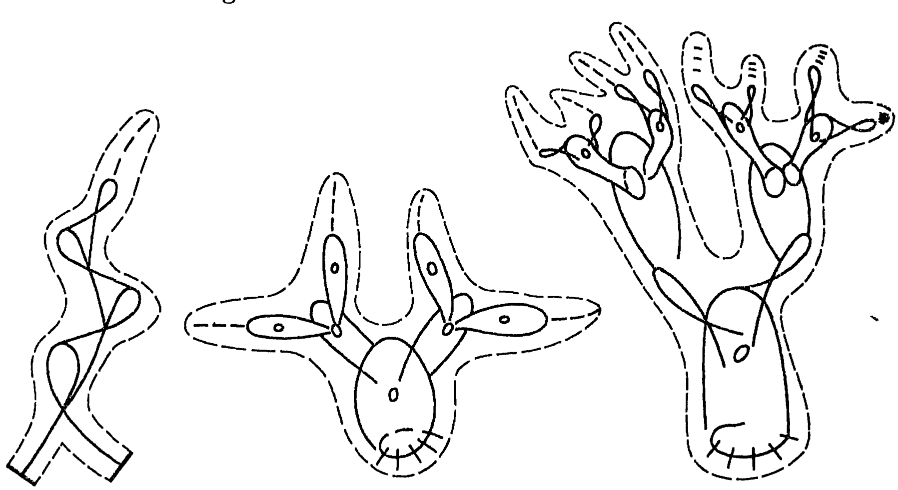
THEOREM 2. Wh_n *is topologically slice for* $n \geqslant 4$.

As in [5] Wh_0 is the Hopf line and Wh_{k+1} is formed from Wh_k by replacing one component of Wh_k by its $\pm$ untwisted double. By using reimbedding theorems, the cases $n \geqslant 5$ are quickly reduced to the simply connected theory. In the case $n = 4$ a surgery problem with $\pi_1 = \boldsymbol{Z} * \boldsymbol{Z}$ is solved by a special argument [9] (which does not appear to work for ramified doubling!). The disk theorem for $\pi_1 M \cong \boldsymbol{Z} * \boldsymbol{Z}$ would, through a strengthening of Corollary 2, imply Theorem 2 for $n \geqslant 3$. For $n \leqslant 2$ known link invariants show that Wh_n is not slice.

Note. It is interesting that the disk theorem, without any restriction to "good groups" (i.e., the disk conjecture), is actually equivalent to the conjunction of Corollary 1 without a group restriction and an unrestricted Corollary 2. This is because the disk conjecture can be translated into a link slice problem in which one must slice a link so that the slice restricted to a certain sublink is unknotted. Surgery for free groups (and the 4-dimensional Poincaré conjecture) would construct the slicing and the s-cobordism theorem for free groups would then be used to verify the required unknottedness.

The topological category is an open invitation to infinite constructions. The proof of the disk theorem depends on maintaining geometric control of the limit set of an infinite 2-complex which is located in the manifold M. Before becoming specific to the proof, I would like to discuss two infinite 2-complexes, their large, but finite, stages, and a hybrid of the two. The 2-complexes to be useful must come with prescribed 4-dimensional thickenings (= open regular neighborhoods of some imbedding in a 4-manifold) and in the infinite case the names *Casson handle* and *infinite grope*[4] for these thickenings seem to be established. The hybrid thickening, the most useful of all (when it can be found!), we will call an *infinite cope*.

The fundamental building block of a Casson handle is a kinky handle [7] which is simply a closed thickening of a disk with finitely many normal self-crossings. The fundamental unit of an infinite grope is a once punctured oriented surface crossed with a two-disk. Infinite copes are built by perpetually alternating one or more layers of surfaces with an immersed disk stage.



(Casson handle, ∂) (Grope, ∂) (Cope, ∂)

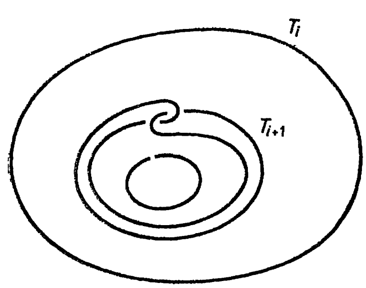
[4] The term "grope" appears in Jim Cannon's work on the double suspension conjecture.

The thickening may be specified[5] by postulating a homeomorphism of pairs (4-manifold, 2-spine) to the "obvious" standard examples of Casson handles, gropes and copes contained in the open 2-handle $\dot{H} = (D^2 \times \dot{D}^2, \partial D^2 \times \dot{D}^2)$.

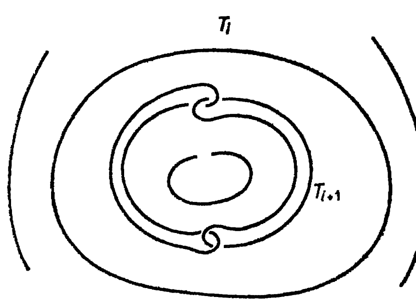
The 2-spines of the standard examples limit to a nice tame Cantor set (or closed subset of same) contained in $\dot{H}$. The positive frontiers Fr^+ [$=$ closure (closure thickening $-$ thickening)] can be analyzed as in [6]. The results are:

$$\begin{aligned} \mathrm{Fr}^+(\text{Casson handle}) &= D^2 \times S^1/\mathrm{Wh}, \\ \mathrm{Fr}^+(\text{grope}) &= D^2 \times S^1/\mathrm{Bing} \cong D^2 \times S^1, \\ \mathrm{Fr}^+(\text{cope}) &= D^2 \times S^1/\text{Alternating Bing–Wh} \cong D^2 \times S^1 \\ &\quad \text{(see footnote 6)}, \end{aligned}$$

where Wh (resp. Bing) denotes the monotone decomposition defined by the intersection of (disjoint unions of) solid tori $\bigcap_{i=1}^{\infty} T_i$ where $T_{i+1} \subset T_i$ is patterned on the untwisted, but possibly ramified, Whitehead (resp. Bing) link. $T_{i+1} \subset T_i$ is called "Whitehead doubling" (resp. "Bing doubling").



Whitehead doubling



Bing doubling

For the positive frontier, Fr^+(cope), the decomposition is defined by alternating finitely many iterations of (ramified) Bing doubling with a single (ramified) Whitehead doubling. Standard shrinking arguments show that Fr^+(Casson handle) is a manifold factor (that is, it becomes a manifold upon crossing with the real line), but not a manifold, whereas Fr^+(grope) and Fr^+(cope)[6] are manifolds (actually solid tori). Intuitively,

---

[5] This amounts to specifying "zero framings" in the attaching of successive stages (see [7]).

[6] I thank the participants of the Santa Barbara Conference on 4-manifolds (Aug. 83) and in particular Ric Ancel for pointing out that Fr^+(cope) will only be a manifold

the singularity is so finely parced by Bing doubling in the last two cases that it simply disappears.

When $\pi_1 \cong 0$, a Casson handle with first stage regular homotopic to j may be constructed directly from the smooth-category version of the disk theorem hypothesis. The main theorem of [7] and the simply connected prototype for 4-dimensional disk theorems was:

THEOREM 3. *Any Casson handle,* CH, *is homeomorphic to the standard open handle* $\dot{H}$.

The proof begins by using reimbedding theorems to laminate CH with copies (a Cantor set's worth) of $\mathrm{Fr}^+(\mathrm{CH})$. A similar lamination (called "design" in [7]) of $\dot{H}$ is constructed. Then some material ($\text{gaps}^+ \subset \mathrm{CH}$, $\text{holes}^+ \subset \dot{H}$) containing the spaces between the leaves of the lamination is crushed out to produce a common quotient Q. Two separate shrinking arguments show that: (1) $\dot{H}$ is homeomorphic to Q (this is Robert Edwards' shrink); and (2) CH is homeomorphic to Q.

Copes can easily be imbedded wherever (or within!) Casson handles are imbedded so it is just as useful to prove:

THEOREM 3′. *Any infinite cope,* C, *is homeomorphic to* $\dot{H}$.

The proof has the same pattern as the Casson handle case but now the laminations, $\mathrm{Fr}^+(C)$, are (undecomposed!) solid tori leading to a simpler design. The homeomorphism of $\dot{H}$ to Q is now much easier to find; it is only necessary to shrink a single countable-null star-like decomposition. This simplification came from discussions with R. Ancel and R. Edwards.

On the other hand, the infinite grope is not simply connected, and, a fortiori, not homeomorphic to $\dot{H}$. To be useful, null homotopies should be attached to produce the notion of capped grope, discussed below.

The pedagogical point is that the infinite cope, combining the best properties of Casson handle and grope, is probably the fastest road into the simply connected theory, and its close relative the capped grope the correct approach to non-simply connected problems.

The previous two theorems are starting points rather than conclusions for the theory. Compare them with Theorem 4.

THEOREM 4. *Any (topological) 4-manifold* $(X, \partial D^k \times \dot{D}^{4-k})$ *which is proper-homotopy equivalent as a pair to the open handle* $(D^k \times \dot{D}^{4-k}, \partial D^k \times \dot{D}^{4-k})$, $k = 0, 1, 2, 3$, *or* 4, *is homeomorphic to the open* k*-handle.*

---

if the number of surface stages between the nth and $(n+1)$st layer of disks is allowed to grow fairly quickly with n, say as 4^n. This feature should be incorporated into the definition of an infinite cope.

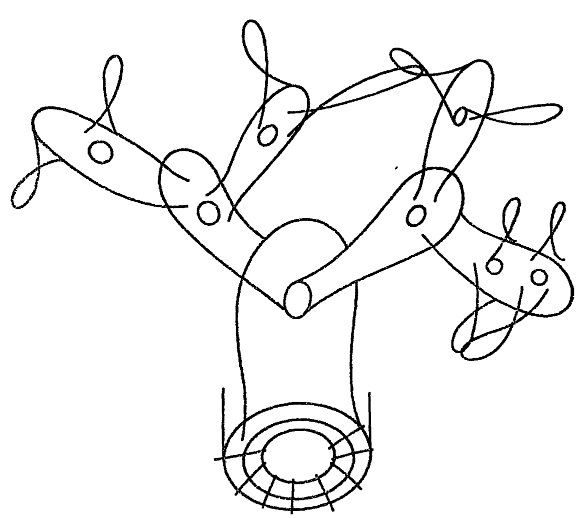
This is a consequence of a proper s-cobordism theorem (compare with Corollary 1). The difficult case is $k = 2$ where the fundamental group of the end is nontrivial. It is stably: $\boldsymbol{Z} \xrightarrow{\cong} \boldsymbol{Z} \xrightarrow{\cong} \boldsymbol{Z} \ldots$, so the disk theorem for $\pi_1 \cong \boldsymbol{Z}$ must be used.

DEFINITION. A *capped grope* $G_{n,1}$ is a compact 4-manifold with a preferred solid torus $\partial^- G_{n,1}$ called the *attaching region* contained in its boundary. Let $\mathrm{SP}_{n+1}(B)$ be the first $n+1$ surface levels in the spine of an infinite grope G, $\mathrm{SP}(G)$, and let i: $\mathcal{N}(\mathrm{SP}_{n+1}(G)) \hookrightarrow M^4$ be an imbedding of a neighborhood of $\mathrm{SP}_{n+1}(G)$ in G into a 4-manifold M.

Let $\{h_j\}$ be a collection of normally immersed null homotopies ("caps") for a simplectic basis of the surfaces at the top, $n+1$-level, of $\mathrm{SP}_{n+1}(G)$. The h_j's are not necessarily disjoint but are required to meet $\mathrm{SP}_{n+1}(G)$ normally and only at their initial time. Anything homeomorphic to closed regular neighborhood $\mathcal{N}[i(\mathrm{SP}_{n+1}) \cup \text{image}\, h_j\text{'s}]$ is a capped grope, $G_{n,1}$.

Addendum. The caps may be assumed to attach with "zero framing". Again this can be defined by reference to the visible model imbeddings in $\dot{H}$. This is without loss since one can imbed the more restricted capped grope in an arbitrary one. Let us incorporate this feature into the definition, as it facilitates the "height raising" lemma which follows.

Note. By a similar refinement we could make the different caps disjoint. We do not do this since disjointness does not propagate with the required efficiency in the later constructions.



Spine and attaching region of a $G_{2,1}$

Like Casson handles, infinite gropes, and infinite copes, a capped grope is a (this time compact) subset of a 2-handle. It is to be investigated because when one is searching for a Whitney disk (read 2-handle) in the typical non-simply connected settings, this is what can be found. Specifically, to prove the disk theorem, one must find Whitney disks to cancel all pairs of double points (p_i, p_i') of j. (Actually the j must be replaced with an immersion j' with even more pairs of double points before the cancellation process can be started.) A capped grope $G_{n,1}$ (attached with a framing suitable for the Whitney trick and with arbitrarily large "grope height" [= first index]) may be found in the place of each Whitney disk.[7]

It is not difficult to show:

FACT. *The disk theorem without fundamental group restriction is equivalent to the existence of a topological 2-handle relatively imbedded in every capped grope $G_{n,1}$, $n \geqslant 1$.*

THEOREM 5. *In any twice capped grope $G_{n,2}$, $n \geqslant 1$, there is a relatively imbedded topological 2-handle.*

A twice capped grope may be defined as a $G_{n,2} = G_{n,1} \cup$ kinky handles where the attaching circles of the kinky handles avoid $\partial^- G_{n,1}$ and normally generate $\pi_1(G_{n,1})$. (The kinky handles may be plumbed to each other as well as self-plumbed.)

Proof. The proof of the big reimbedding theorem [7] adapts to (and actually simplifies in) the twice capped grope setting. Successive applications eventually produce a grope capped off by (0-framed) 6-stage towers, $G_{n,6}$, relatively imbedded in $G_{n,2}$. Using the existence of topological handles in 6-stage towers one finds the relatively imbedded nest:

$$\text{2-handle} \subset G_n \cup \text{2-handles} \subset G_{n,6} \subset G_{n,2}. \ \square$$

The disk theorem is proved by constructing capped gropes ($G_{K,1}$'s) as substitute Whitney disks on j' and then finding reimbeddings for the

[7] Hint for finding these capped gropes: First find disjoint immersed Whitney disks d_i with paired self-intersections, i.e., $\mu(d_i) = 0$. Tube these paired points together, thereby replacing the immersed disks by surfaces; these will be the bottom stages of the capped gropes. Now if the d_i's were equipped with transverse spheres (= dual spheres = π_1-negligible), the surfaces will have transverse spheres and it will be possible to attach a layer of zero framed, $\mu = 0$, disjoint, π_1-negligible, null homotopies to a simplectic basis on the surfaces. This constructs a $G_{0,1}$. Repeat the cycle to raise the first index.

capped gropes, $G_{k,1} \hookrightarrow G_{K,1}$, $k \ll K$, so that the composition:

$$\pi_1(G_{k,1}) \xrightarrow{\text{inc}_\#} \pi_1(G_{K,1}) \xrightarrow{\text{inc}_\#} \pi_1(M) \qquad (*)$$

is the zero map. This enables a second layer of caps to be added. An application of the previous theorem yields the topological two-handle needed for the topological Whitney trick. This will give the topological regular homotopy claimed in the conclusion of the disk theorem. Reimbedding the capped grope is a delicate problem and the solution depends on the fundamental group of M being in some sense small.

The starting point for reimbedding capped gropes is:

THE GROPE HEIGHT RAISING LEMMA 1. *Given any capped grope $G_{n,1}$, $n > 1$ and any integer $N > 0$, there is a relative imbedding $G_{n2^N,1} \hookrightarrow G_{n,1}$ which is a homeomorphism on the attaching regions.*

I would like to thank Bob Edwards for suggesting a simple combinatorial procedure for raising grope height. This enables one to count (carefully) word length.

Addendum. Given a capped grope $G_{j,1}$, one may choose a base point on the base surface of $\mathrm{SP}(G_{j,1})$. Every double point of the disks contained in $\mathrm{SP}(G_{j,1})$ (i.e., every self-intersection or intersection among the caps) determines an element (up to inverse) in $\pi_1(G_{j,1})$. These elements are the natural free generators for $\pi_1(G_{j,1})$. We claim that under the inclusion produced by the Lemma, $G_{n2^N,1} \hookrightarrow G_{n,1}$, the natural fundamental group generators of $\pi_1(G_{n2^N,1})$ are carried to words of length less than or equal to 7^N in the natural fundamental group generators of $\pi_1(G_{n,1})$.

The reimbedding technique alluded to above may reduce the cardinality of the *loop set* $\{\text{inc}_\#(\text{natural generators } \pi_1(G_{K,1})\} \subset \pi_1(M)$ at the expense of decreasing grope height (recall $k \ll K$). The height raising lemma restores grope height at the expense of enlarging the set. The opposite character of these two techniques leads to a "race" between grope height and loop set cardinality, the outcome depending on $\pi_1(M)$.

Central to grope height raising and reimbedding is the existence of a certain unobvious null homotopy $\bar{h}$. For a detailed discussion of this see Edwards' analysis of Quinn's proof of the annulus conjecture [5]. Here, the essential properties of $\bar{h}$ are described; its construction is treated as a difficult exercise.

A symmetric null homotopy. Let $C = (S \times D^2, \partial S \times D^2) \cup k_1 \cup \ldots \cup k_n \cup \cup k_1' \cup \ldots \cup k_n'$ be a $G_{0,1}$ capped grope (equally a "one stage cope"). The k' and k's are kinky handles [6] (possibly with plumbings between distinct handles) attached by any framings to a simplectic basis $(a_1, \ldots, a_n, a'_1, \ldots \ldots, a'_n)$ of an oriented, once punctured, genus n surfaces.

There are two "obvious" null homotopies h and h' for $\partial S \times 0 \subset C$. These come from surgery of $S \times 0$ either along $(k_1, \ldots, k_n)$ or along $(k'_1, \ldots \ldots, k'_n)$. The natural generators (up to inverse) of $\pi_1(C)$ may be divided into three groups $\{x\text{'s}\} \cup \{x'\text{'s}\} \cup \{y\text{'s}\}$, x's (resp. x''s) arising from self-plumbing of the k's (resp. k''s) and y's arising from plumbings between the collections. The images of h and h' also have natural (up to inverse) free fundamental group generators. These generators map to $\{x\text{'s}\} \subset \pi_1(C)$ [resp. $\{x'\text{'s}\} \subset \pi_1(C)$]. There is a third "unobvious" null homotopy $\bar{h}$ which symmetrically utilizes both sets of kinky handles. The double points of $\bar{h}$ may be arranged[8] so that the set $\mathcal{S}$ of natural generators for $\pi_1(\text{image } \bar{h})$ contain certain "differences" of x's and x''s. Specifically, let X and X' satisfy: $\{x\text{'s}\} \subset X$ and $\{x'\text{'s}\} \subset X'$ and $X \amalg X' = \{x\text{'s}\} \amalg \{x'\text{'s}) \amalg \{y\text{'s}\}$ (and that the ambiguity: element $\equiv$ element^{-1} is resolved in X and X', arbitrarily for x's and x''s and by the assignment to X or X' for y's). Let $X_0 \subset X$ and $X'_0 \subset X'$ be any subsets. Then for these choices $\bar{h}$ can be arranged so that

$$\mathcal{S} = \{b^{-1}a \mid a \in X_0 \text{ and } b \in X'_0\} \cup \{X - X_0\} \cup \{X' - X'_0\}. \qquad (**)$$

The last two stages of any $G_{K,1}$ may be replaced by thickened null homotopies modelled on image$(\bar{h}) \subset C$. This gives a "reimbedding" $G_{K-1,1} \subset G_{K,1}$ with a new (and strictly smaller for certain choices of X_0 and X'_0) loop set.

The symmetric null homotopy $\bar{h}$ has a second feature. It is possible to change the spine $\mathrm{SP}(C)$ by Casson's finger moves between the caps to $\widetilde{\mathrm{SP}}(C)$ so that $\widetilde{\mathrm{SP}}(C)$ and image$(\bar{h})$ are disjoint.

The second property of $\bar{h}$ may be iterated to produce arbitrarily many disjoint null homotopies for parallel copies of $\partial S \times 0$ in C.[9] In the proof of the annulus conjecture and, for example, in the proof of Theorem 5, this can be used to produce a large supply of dual spheres (to some 2-complex) starting with one capped surface dual to the 2-complex.

Proof of Lemma 1. The first step is to do finger moves to the caps in $G_{n,1}$ to create a transverse collection of transverse spheres to the caps (these are framed immersed spheres with $\mu = 0$ meeting the caps "geometrically-δ_{ij}"). To produce the collection, turn the (visible) tori trans-

---

[8] This arrangement was discovered using Kirby's link calculus although it can be explained in several ways.

[9] This key fact is foreshadowed in John Milnor's senior thesis on link homotopy [12] which contains the logical forerunner of the Casson finger move.

verse to the caps into spheres by surgery. The surgery is done along null homotopies which avoid the changed spine $\widetilde{\mathrm{SP}}(G_{n,1})$. (Note that we have used the second property of the null homotopy $\bar{h}$.) Let $\tilde{G}_{n,1} \subset G_{n,1}$ be a neighborhood of the changed spine.

Second, regard $\mathrm{SP}(G_{n,1})$ $[\mathrm{SP}(\tilde{G}_{n,1})]$, above its base stage, as the union (not necessarily disjoint) of a right branch $R[\tilde{R}]$ and a left branch $L[\tilde{L}]$. $(R \cup L)$ can be perturbed to be disjoint from $(\tilde{R} \cup \tilde{L})$ except for crossings at the caps. By tubing into copies of the transverse collection created in step one we form R' and L' with $(R' \cup L') \cap (\tilde{R} \cup \tilde{L}) = \emptyset$.

Third, pipe down the double points of the caps of $\tilde{R}$ to the base stage. There the double points are removed by splicing many copies of L' into each cap. (In fact, 2^{n+1} copies of L' are used to remove each double point.)

Fourth, pipe down the double points of the caps of $\tilde{L}$ and the intersection points between the caps of $\tilde{L}$ and the caps of $\tilde{R}$ to the base stage. Remove these by splicing in copies of R' whose surface stages are so close to $\tilde{R}$ as to be disjoint from the pipes constructed in step three.

A neighborhood of the original surface stages and the modified caps, $\mathcal{N}(G_{n,0} \cup$ modified caps) is now a $G_{2n,1}$ contained in the original $G_{n,1}$. The lemma follows by repeating steps one through four.

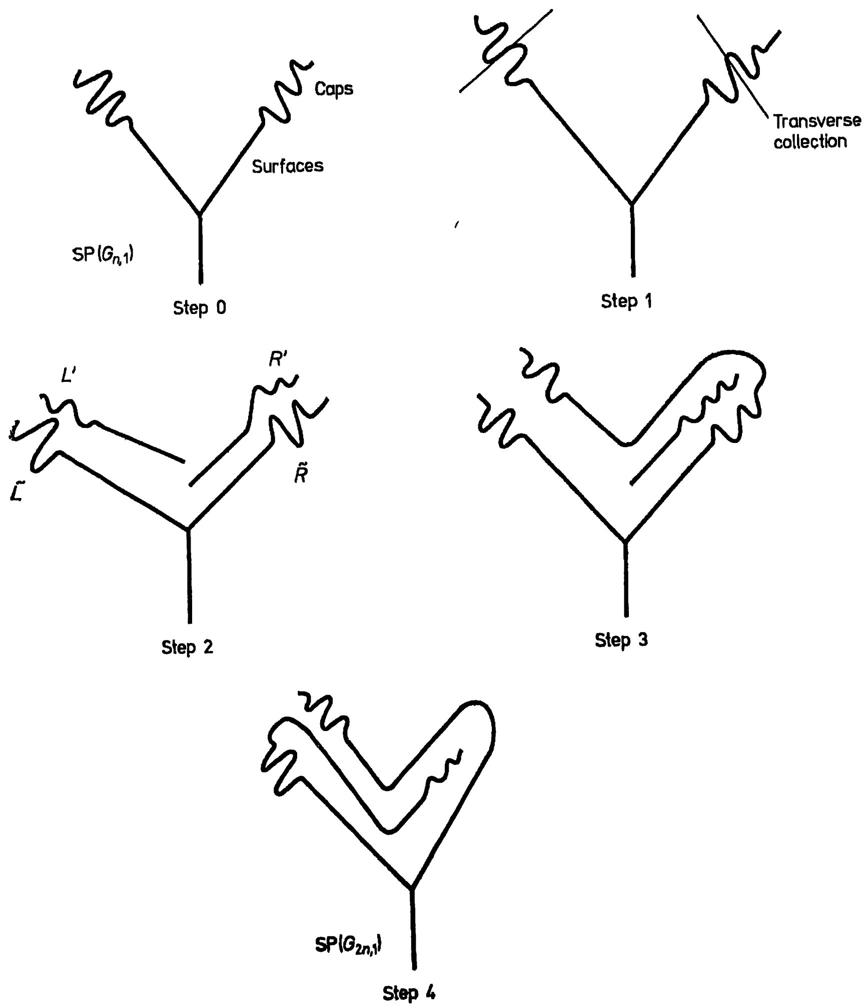
The steps of the proof may be diagrammed as shown on p. 659.

Although checking the addendum requires careful thought, we can quickly count to seven. Word length for loops in the transverse spheres is one, tubing (Step 2) conjugates by the (length one) intersection point of cap and sphere. So, for example, L' has length $1+1+1 = 3$. By looking at the finger moves needed to produce the transverse collection, we see that word length in $\tilde{R}$ is two. Forming the new right branch (Step 3) again involves conjugation, so the final word length is $2+3+2 = 7$.

For the proof of the disk theorem we must arrange $\mathrm{inc}_{\#}\colon \pi_1(G_{k,1}) \to \pi_1(M)$ to be zero (see line(*)). The tools are Lemma 1 and line (**) which describes the possible benefits to the loop set available at the price of sacrificing one surface stage.

Suppose $\pi_1(M)$ is finite. Apply Lemma 1 until $n2^N \gg \mathrm{order}[\pi_1(M)]$. By choosing $X_0 = Y_0$ to be singletons $\{g\}$, $g \neq e \in \pi_1(M)$, we see that the loop set can be reduced by one element by reducing grope height by one. In this way one can reduce the loop set to $\{e\}$ for some reimbedded capped grope $G_{h,1}$ with $h \geqslant 1$. Now $G_{h,1}$ can be extended to a twice capped grope $G_{n,2}$, completing the proof of the disk theorem in the case of finite fundamental groups.

Suppose $\pi_1(M) \cong \mathbf{Z}$. We exploit the ambiguity — element $\equiv$ element^{-1} — to always replace an integer pair $\{n, -n\}$ with its absolute



value, $|n|$. Thus the loop set becomes a collection of nonnegative integers contained, say, in the interval $[0, 2k+1]$. If we set $X_0 = X \cap \{k+1, k+2, \ldots, 2k+1\}$ and $Y_0 = Y \cap \{k+1, k+2, \ldots, 2k+1\}$ then the new loop set will be contained in $[0, k]$. Thus the diameter of the loop set can be at least halved at the cost of reducing grope height by one. So the number of grope stages which must be expended to bring the loop set down to $\{e\}$ is no more than $\lceil \log_2(\text{diameter}(\text{loop set})) + 1 \rceil$.[10] Suppose loop set $(G_{n,1})$ has diameter d and Lemma 1 is applied for a given value of N. By the

[10] Bracket denotes "least integer greater than".

Addendum, diameter$\left(\text{loop set}(G_{n2^N,1})\right) \leqslant d(\text{maximum word length}) = d7^N$. To bring the loop set to $\{e\}$ we need sacrifice no more than $\log_2(d7^N)+2$ stages. Recall that n and d are constant, so for sufficiently large N,

$$n2^N \geqslant \log_2(d7^N)+2.$$

Consequently, the desired $G_{h,1}$ with trivial loop set can be constructed to lie in a nest:

$$G_{h,1} \subset G_{n2^N,1} \subset G_{n,1}.$$

As in the finite group case, extend $G_{h,1}$ to $G_{h,2}$ to complete the proof.

Notice the "loop set" can be defined in any homomorphic image of $\pi_1(M)$. The previous two arguments can be reinterpreted as procedures for bringing the loop set into the kernel of any homomorphism $\pi_1(M) \to H$ where H is a finite group or the integers. Thus the groups for which a sequence of the above procedures will reduce any loop set to $\{e\}$ is closed under group extensions. It is also closed under infinite nested union. Thus the disk theorem is extended to infra-solvable fundamental groups.

Applications

In addition to the general Corollaries 1 and 2, some applications of the disk theorem deserve mention. In some cases the 4-dimensional information completes, or nearly completes, a dimension-free pattern in the topological category. In each statement, however, the new information is in just one dimension.

THEOREM 6. *A flat topological knot $S^{n-2} \hookrightarrow S^n$ is unknotted iff the complement $S^n \setminus S^{n-2}$ has the homotopy type of a circle. (Having the homotopy type of a circle is equivalent to $\pi_1(S^4 \setminus S^2) \cong \mathbf{Z}$ in the case $n = 4$.)*

Remarks. This is a "Dehn's Theorem" in dimension $n = 3$, and a result of Stallings [17] for $n \geqslant 5$. The proof is by the 5-dimensional-topological-S-cobordism theorem for $\pi_1 \cong \mathbf{Z}$ ($\mathbf{Z}$-5DTSCT).

THEOREM 7. *Let $k: S^1 \hookrightarrow S^3$ be a classical (tame) knot. The knot k is the boundary of a flat topological disk K: $D^2 \hookrightarrow B^4$ with $\pi_1\left(B^4 \setminus K(D^2)\right) \cong \mathbf{Z}$ iff the Alexander polynomial of k is trivial, $\Delta_k(t) = 1$ ["flat" means the imbedding K extends to an imbedding $\bar{K}$: $(D^2 \times \mathbf{R}^2, \partial D^2 \times \mathbf{R}^2) \hookrightarrow (B^4, S^3)$].*

Remarks. By Milnor duality [13] a slice with fundamental group complement $\cong \boldsymbol{Z}$ implies $\tilde{H}_*(S^3\setminus k;\, \boldsymbol{R}[\boldsymbol{Z}]) \cong 0$. Since $\Delta_k(t)$ is by definition the determinant of the relation matrix for the module when $* = 1$, $\Delta_k(t)$ is a unit, and by conversion normalized to one. Conversely, Crowell [3] has shown $\Delta_k(t) = 1 \Leftrightarrow \tilde{H}_*(S^3\setminus k;\, \boldsymbol{Z}[\boldsymbol{Z}]) \cong 0$. This allows one to construct a surgery problem (see [8]) over $\pi_1 \cong \boldsymbol{Z}$ with vanishing obstruction in $L_4(\boldsymbol{Z}) \cong \boldsymbol{Z}$. Apply Corollary 2 with $\pi_1 \cong \boldsymbol{Z}$ to solve the surgery problem and thereby construct a candidate for the closed slice complement. Glue in the slice and use the 4-dimensional Poincaré conjecture to recognize the results as homeomorphic to B^4.

THEOREM 8. *Every n-dimensional manifold homotopy equivalent to an n-torus $S^1 \times \ldots \times S^1$ (n copies) or an n-sphere S^n, for $n \neq 3$, is homeomorphic to an n-torus or n-sphere, respectively.*

Remarks. The proof in the torus case requires computing homotopy classes of maps $[T^4, G/\mathrm{top}]$, the action of $L_5(\boldsymbol{Z}^4)$ on $[\Sigma T^4, G/\mathrm{top}]$ and an application of $\boldsymbol{Z}^4$-5DTSCT. For the sphere case see [7].

THEOREM 9. *Any smooth 4-manifold homotopy equivalent to the real projective 4-space $\boldsymbol{RP}^4$ is homeomorphic to the same.*

Remarks. Cappell and Shoneson [1] have computed that any fake $\boldsymbol{RP}^4$ with Kirby-Siebenmann obstruction equal to zero is topologically s-cobordant to $\boldsymbol{RP}^4$; apply $\boldsymbol{Z}_2$-5DTSCT.

THEOREM 10. *Let $S^{n-2} \subset M^n$ be a complete topological codimension two submanifold. S^{n-2} has a neighborhood homeomorphic as a pair to a vector bundle over S^{n-2} iff S^{n-2} is locally homotopically unknotted (i.e., $\forall s \in S^{n-2}$ and $\forall$ neighborhoods U_s $\exists V_s \subset U_s$ s.t. any loop $\gamma \in V_s\setminus S^{n-2}$ which has local linking number zero with S^{n-2} is null homotopic in $U_s\setminus S^{n-2}$.*

Remarks. Classically, for $n \neq 4$, this is approached by engulfing to establish local flatness [14] and then applying a torus trick argument [11]. (These papers still leave a small gap which was filled by [2] and slightly later by [16].) The 4-dimensional case is proved by destabilizing inc $\times$id: $S^2 \times \boldsymbol{R} \hookrightarrow M^4 \times \boldsymbol{R}$ using Quinn's controlled 5-dimensional h-cobordism theorem. In his paper [15] he proves the analogous result for manifolds $N^3 \hookrightarrow M^4$. His proof does not apply directly because the local homotopy data give no control in the direction linking the surface. (That is, we do not have a "δ-1-connected ε-h-cobordism".) One must accept this lack of control and still construct Whitney disks in the middle level (of the h-cobordism to the standard object). This requires finding 2-handles given once capped gropes. The effective fundamental group for this problem is the integers, i.e., local linking numbers with the submanifold S^2.

Metatheorem 1. Using Corollaries 1 and 2 for finite groups, it will be possible to obtain a lot of information on the topological classification of finite groups acting on compact 1-connected 4-manifolds.

Metatheorem 2. Versions of the surgery and s-cobordism theorems for "singular" 4-manifolds can be obtained. The point singularity will be non-ANR but still have the local integral homology of a manifold. The idea is to exploit the fact that general surgery [s-cobordism theorem] can be reduced to the $\pi_1 \cong$ free case. The surgery [s-cobordism] theorem is known for solvable groups, and free groups are residually solvable, so it is not surprising that there are limiting constructions which bridge the gap. Unfortunately, the present constructions all lead to singularities of the type described.

References

[1] Cappell S. and Shoneson J., Some New 4-Manifolds, *Ann. of Math.* **104** (1976) pp. 61–72.

[2] Chapman T. A., Locally Homotopically Unknotted Embeddings of Manifolds in Codimension Two Are Locally Flat, *Topology* **18** (1979), pp. 339–348.

[3] Crowell R. H., The Group G'/G'' of a Knot Group G, *Duke Math. J.* **30** (1963), pp. 349–354.

[4] Donaldson S., An Application of Gauge Theory to 4-Dimensional Topology, *J. Differential Geometry* **18** (1983), pp. 279–319.

[5] Edwards R., Article in: *Proceedings of the AMS Conference on 4-Manifolds*, Summer 1982, Durham, New Hampshire (to appear).

[6] Freedman M. H., A Fake $S^3 \times \boldsymbol{R}$, *Ann. of Math.* **110** (1979), pp. 177–201.

[7] Freedman M. H., The Topology of 4-Dimensional Manifolds, *J. Differential Geometry* **17** (1982), pp. 357–453.

[8] Freedman M. H., A Surgery Sequence in Dimension Four; the Relations with Knot Concordance, *Invent. Math.* **68** (1982), pp. 195–226.

[9] Freedman M. H., Slicing Wh_4, Austin Texas lecture notes (Nov. 1982, to appear).

[10] Kirby R. and Siebenmann L., *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Annals Math. Studies **88**, Princeton Univ. Press, Princeton, 1977.

[11] Kirby R. and Siebenmann L., Normal Bundles for Codimension 2 Locally Flat Imbeddings. In: *Conference on Geometric Topology, Park City, Utah* (1974), Lecture Notes in Math. 438, Springer-Verlag, pp. 310–324.

[12] Milnor J., Link Groups, *Ann. of Math.* **59** (1954), pp. 177–195.

[13] Milnor J., A Duality Theorem for Reidemeister Torsion, *Ann. of Math.* **76** (1962), pp. 137–147.

[14] Price T. and Seebeck C., A Codimension Two Taming Theorem. In: *Conference on Geometric Topology, Park City, Utah* (1974), Lecture Notes in Math. 438, Springer-Verlag, pp. 371–394.
[15] Quinn F., Ends of Maps III, *J. Differential Geometry* **17** (1982), pp. 503–521.
[16] Quinn F., Ends of Maps II, *Invent. Math.* **68** (1982), pp. 353–424.
[17] Stallings J., On Topologically Unknotted Spheres, *Ann. of Math.* **77** (1963), pp. 490–503.
[18] Whitney H., The Self-Intersections of a Smooth N-Manifold in $2N$-Space, *Ann. of Math.* **45** (1945), pp. 220–246.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT SAN DIEGO, LA JOLLA, CALIFORNIA 92093, USA

**Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa**

STEVEN P. KERCKHOFF*

The Geometry of Teichmüller Space

In this paper we will discuss some recently discovered geometric properties of Teichmüller space. We will primarily think of Teichmüller space as the space of geometric (flat or hyperbolic) structures on a surface of genus g, $g \geqslant 1$, and concentrate on deformations of these structures. The flat case has been well understood for many years. This is due, at least in part, to the homogeneity of the Teichmüller space in this case; it is naturally isometric to 2-dimensional hyperbolic space, $\boldsymbol{H}^2$. The goal of this paper is to outline an approach to the genus 1 case which generalizes fairly well to higher genus. The key links are between the geodesic length function, geodesic laminations (foliations) and 1-parameter families of deformations.

In Section I our point of view is developed in the familiar genus 1 setting. Section II discusses the higher genus case. A proof of the Nielsen realization problem is sketched here, the original motivation for the development of this material. The proof is slightly different from the published version, yielding somewhat finer results. Finally, in Section III, we outline some further developments and discuss some related open questions.

Unfortunately, the perspective here is necessarily limited. Almost no mention is made of the interplay between quasi-conformal mappings and Teichmüller theory. The reader is referred to the survey article of Bers [1] and its extensive bibliography for a thorough description. There has also been a great deal of more recent work which must go unmentioned or only briefly acknowledged. The author has been particularly influenced by Thurston's work as well as that of Masur, Wolpert, Jørgensen, Bers, and others. The reader desiring a broader perspective is encouraged to supplement this note with the works of those authors listed in the bibliography.

* This research was partially supported by an N.S.F. grant MCS82-03806.

I

The *Teichmüller space* of genus g, T_g is the space of conformal structures on a surface of genus g where two are considered equivalent if there is a conformal map between them which is isotopic to the identity. Without the restriction that the map be isotopic to the identity, we would get the moduli space, M_g, of genus g which is thus the quotient space of T_g under the action of the group of conformal self-maps. This group will be called the *modular group*, Mod_g, because for $g = 1$ it is the classical modular group. Mod_g is also the group of isotopy classes of homeomorphisms of our surface, M, and is known to be isomorphic to the group of outer automorphisms of $\pi_1 M$.

It follows from the uniformization theorem that for $g \geqslant 2$, T_g is also the space of hyperbolic structures (metrics of constant curvature -1) on a fixed surface of genus g where two are equivalent if they are isometric by an isometry isotopic to the identity. For $g = 0$, T_g is the space of spherical structures on S^2, but since such a structure is unique, we ignore this case. For $g = 1$, T_g is the space of equivalence classes of flat metrics of area 1 on the torus. Although our main concern will be the case $g \geqslant 2$, our theme is the extent to which properties of the genus 1 case generalize to higher genus. Thus we begin with a discussion of that case.

A flat metric on the torus is given by a two-dimensional lattice in $\boldsymbol{R}^2$ which, in turn, is given by an ordered pair of linearly independent vectors. We identify this pair of vectors with a matrix which, because of the area convention, is in $\mathrm{SL}(2, \boldsymbol{R})$. Two such lattices are equivalent iff they differ by a rotation so we find that for $g = 1$ T_g is identified with $\mathrm{SL}(2, \boldsymbol{R})/\mathrm{SO}(2)$ which, with its usual left invariant metric, is isometric to $\boldsymbol{H}^2$. Given this nice identification of T_1 with hyperbolic space, it is natural to ask what various geometric objects in $\boldsymbol{H}^2$ correspond to in T_1.

We begin with square torus, which is identified with the identity matrix, and consider the family of tori corresponding to the matrices $\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$. This traces out a geodesic in hyperbolic space which we assume goes from 0 to ∞ in the upper half space model. The corresponding family of tori are those with a rectangular lattice and can be gotten from the square torus by an affine deformation which lifts to $\boldsymbol{R}^2$ as the map $(x, y) \mapsto (e^t x, e^{-t} y)$. This map preserves two foliations on $\boldsymbol{R}^2$, one by horizontal lines, the other by vertical lines. These foliations descend to two orthogonal foliations of the tori by closed geodesics whose homotopy classes we denote by α and β. Thus the deformations of the flat structure along this curve in T_1 can be described as squeezing the metric along β and stretching it along a for $t > 0$ and vice versa for $t < 0$. As $t \to \infty$, the length of β goes to zero; the length of a goes to zero as $t \to -\infty$. We identify the points 0, ∞ on the boundary of $\boldsymbol{H}^2$ with β, a, respectively.

If we multiply $\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$ on the left by any element of SO(2) we get a new geodesic in $\boldsymbol{H}^2$ whose deformations can similarly be described in $\boldsymbol{R}^2$ by an affine stretching and squeezing along two orthogonal sets of parallel lines. These descend to the torus as geodesic foliations; for a dense set of elements of SO(2) the geodesics are simple closed curves, but in the other cases the leaves of the foliation are infinite simple geodesics. We can define a transverse Borel measure μ whose support is the whole foliation and which is proportional to the Euclidean distance between leaves. (Hence it is unique up to scalar multiplication.) Given this measure we can define the intersection number $i(\mu, \cdot)$ of ac urve (or arc) to be infimum of the integral of the measure over the curves isotopic to the given one. (In the case of an arc the endpoints are required to be fixed.) The infimum is always realized when the curve is a geodesic.

Take the product of this measure with the Lebesgue measure induced by distance along leaves of the foliation and define the length of the foliation (with transverse measure) to be the total mass of this product measure. When the foliation consists of parallel simple closed curves and the transverse measure is normalized so that the intersection number with an arc hitting each leaf once is 1, this length coincides with the usual length of the geodesic curves.

Fixing a choice of the transverse measure of each geodesic foliation on the square torus, it is apparent that during the 1-parameter family of deformations, the length of the foliation being squeezed is going to zero and all the others are going to ∞. We identify the endpoint of the geodesic with the foliation whose length is going to zero so that the points at infinity (and directions in the tangent space of T_1) are in 1–1 correspondence with geodesic measured foliations (up to scalar multiple) on the square torus.

All other geodesics in $\boldsymbol{H}^2$ are conjugate to those through the square torus and the corresponding deformations on the tori are similarly defined. There are two orthogonal (in the particular flat metric) geodesic foliations together with a one-parameter family of affine maps squeezing along one and stretching along the other. It is not hard to see that if, along two distinct geodesics, the foliations whose lengths are going to zero consist of simple closed curves, isotopic on the underlying topological surface, then the endpoints of the two geodesics are equal. In fact it is true in general that the endpoints of two geodesics are equal iff the contracting foliations are isotopic. Thus the identification of a point at infinity with an isotopy class of a measured geodesic foliation is independent of the surface with which one begins.

A second basic object in the geometry of $\boldsymbol{H}^2$ is a horocycle, a curve of constant geodesic curvature $+1$. All horocycles are images under an isometry of the line $y = 1$ in the upper half-space model. They are circles tangent to the circle at infinity, and are orthogonal to all geodesics with an endpoint at that point at infinity.

The one-parameter family of matrices corresponding to the horocycle $y = 1$ is $\begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}$. On the square torus this deformation is a shearing map which preserves the simple closed geodesics which are the projections of horizontal lines in $\boldsymbol{R}^2$. The geodesics are isotopic to a, one of the two geodesics which determined the original stretch deformations. The amount of shearing at time t along a transverse geodesic on the surface is equal to t times the intersection number of the geodesic with the geodesic foliation by parallel copies of a (with suitably normalized measure). The same path can be described by cutting along a single closed geodesic, twisting to the left distance proportional to t and re-gluing. It is harder to describe this when the leaves are non-compact; however, this is the form that generalizes to higher genus.

Conjugate horocyclic deformations can similarly be described on the torus by a shearing map which preserves the leaves of a geodesic measured foliation μ where the amount of shearing (always to the left) along a transverse geodesic arc A at time t equals $ti(\mu, A)$. The length of μ is preserved under this horocyclic deformation so the horocycle is contained in a level set of the length function. Because the dimension of the level set is 1, the path is, in fact, the entire level set if we allow all $t \in \boldsymbol{R}$.

One way to generalize some of this structure to higher genus comes from the theory of quasi-conformal mappings. Given a Riemann surface M and a quadratic differential $\theta(z)$ on M, we can define a one-parameter family of deformations of the conformal structure as follows: Away from the singularities of $\theta(z)$ we can find a local co-ordinate ω so that θ equals $d\omega^2$. The condition $\theta > 0$ (resp., < 0) determines a *horizontal* (resp., *vertical*) foliation (with singularities) with a transverse measure derived from the metric $ds = |\theta|^{1/2}$. These are the lines $x = \text{const}$ (resp., $y = \text{const}$) in the local co-ordinate $\omega = x+iy$. With respect to this co-ordinate, the deformation is simply the affine map $(x, y) \mapsto (e^t x, e^{-t} y)$. It is not hard to check that this change of conformal structure is globally well-defined and can be extended over the singularities of θ.

The curves in T_g so defined are called *Teichmüller rays*. Teichmüller showed that these rays are the solution to an extremal problem for quasi-conformal mappings and proved that there is a unique such ray from M to any other point in T_g (see [3], e.g., for a modern proof). Thus T_g is homeomorphic to a $(6g-6)$-dimensional ball (the dimension of the space of quadratic differentials) which can be compactified by a $(6g-7)$-dimensional sphere corresponding to the endpoints of the rays from M. These rays are geodesics in the Teichmüller metric on T_g ([14]) so this is a metric without conjugate points. Moreover, Mod_g acts isometrically on T_g in this metric.

Unfortunately, at this point the suggestive analogy with hyperbolic space begins to unravel. In contrast, the Teichmüller metric does not have non-positive curvature ([16]) and the compactification by endpoints depends on the choice of base point M for the rays ([8]). Thus neither the local geometry nor the asymptotic geometry of this metric quite parallels that of hyperbolic space.

Although not all rays behave asymptotically like those in hyperbolic space, it follows from the proofs in [8] that the bad boundary behavior of the compactification does not occur for rays whose horizontal foliations are uniquely ergodic (see [17], e.g., for a definition and [19] for a closely related result). Masur showed that this set has full measure in the tangent space of every point ([18], see also [20], [13]). Thus, from the point of view of measure theory, the geodesic structure is very much like that of hyperbolic space. In fact, Masur further proves ([17]) that for this subset of the tangent space, corresponding rays from distinct points in T_g actually approach each other asymptotically. The volume of the moduli space is finite ([18]) so one can apply Hopf's proof for hyperbolic manifolds of finite volume to get the following striking result:

THEOREM 1.1 (Masur [18]). *The geodesic flow of the moduli space in the Teichmüller metric is ergodic.*

The lack of negative curvature for T_g is less reasonably ignored. In particular, because of this problem, it remained unknown whether every finite subgroup of Mod_g acting on T_g has a fixed point. Such a fixed point would be a surface whose conformal structure or hyperbolic structure is symmetric under this finite group. The existence of this fixed point is, in fact, equivalent to the *Nielsen realization problem* that every finite subgroup of Mod_g can be realized as a group of isometries of some hyperbolic surface. For $g = 1$ (and flat metrics) the solution is classical. An easy proof is to look at the orbit in $T_1 = \boldsymbol{H}^2$ of any point under the finite group. It will have a unique barycenter by the convexity of the metric on $\boldsymbol{H}^2$. The barycenter is therefore a fixed point since it is unique.

In the next section we discuss how some new convexity properties of T_g allow us to solve the Nielsen problem in general. The key object is the geodesic length function of simple closed geodesics on hyperbolic surfaces. Finally, in the last section, we describe some recent progress towards a geometry, based on the length function, generalizing the hyperbolic geometry of T_1.

II

If we think of T_g as the space of hyperbolic structures on M, a natural function to consider is the geodesic length, $l_\emptyset$, if a homotopy class of a simple closed curve, $\emptyset$, with respect to the hyperbolic metric. In contrast, extremal length is more central to the geometry of the Teichmüller metric (see, e.g., [8], Theorem 4). In the flat case the level sets of these functions were horocycles; for higher genus the level sets are qualitatively like horospheres. The interiors of the horocycles, "horoballs", are described by the condition $l_\emptyset \leqslant$ const. It is important to note that the horoballs are convex; we shall see that this is the central quality that generalizes to higher genus. Before discussing this generalization, we first see how this property can be used to solve the Nielsen problem in the classical case.

Consider the finite subgroup G of order 2 in Mod_1 which interchanges the curves α and β representing the standard generators of $\pi_1 M$. The horoballs $l_\alpha \leqslant C$, $l_\beta \leqslant C$ intersect in a compact, non-empty set, I_C, for C sufficiently large (see Figure 1a). I_C is G-invariant since the curves $\{\alpha, \beta\}$ are an entire orbit under G. Let $C \to 0$. For C sufficiently small, I_C is empty since the area of the torus ($= 1$ by definition) is less than the product $l_\alpha l_\beta$. Thus, there is a least value, C_0, of C when I_C is non-empty. By convexity I_{C_0} is a point (in fact, by inspection, the point of tangency of the sets $l_\alpha = 1$, $l_\beta = 1$) (Figure 1b). Since it is G-invariant, it is the fixed point we were seeking.

The proof is the same for a general finite subgroup. For example, an orbit for the order 3 subgroup consists of the curves α, β, and γ, where α, β are as before and γ represents the sum in $\pi_1 M$ of the two standard generators. Figures 2a, 2b describe the sets I_C, large C, and I_{C_0}, respectively (T_1 is normalized differently from Figures 1a, 1b for the sake of symmetry), the latter being the fixed point for this subgroup. Note that in this case the fixed point is the unique point where $l_\alpha = l_\beta = l_\gamma$, a necessary condition for a fixed point.

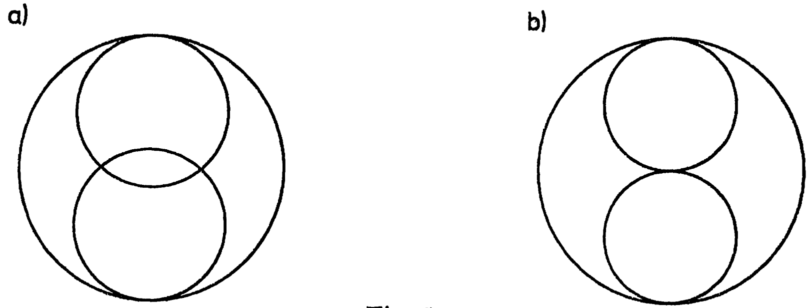


Fig. 1

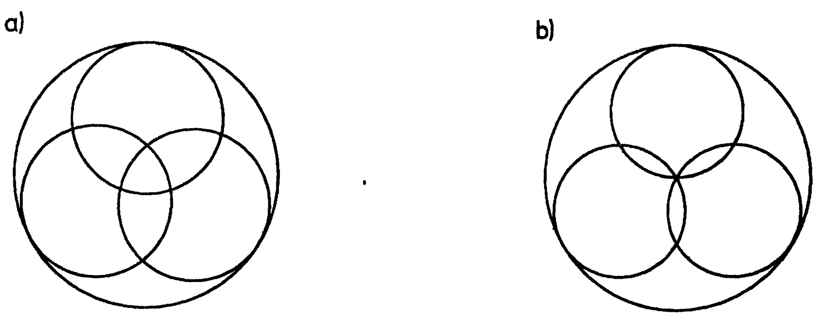


Fig. 2

The horoballs in T_1 are convex in a number of ways, with respect to Euclidean lines, to hyperbolic lines, and to horocycles. The Euclidean lines have no apparent connection to deformations of flat structures, and it is not so clear how to generalize the stretch deformations corresponding to hyperbolic lines in the context of hyperbolic geometry. (More on this in the next section.) However, if we recall that the horocycles were, in many cases, the traces of 1-parameter Dehn twists, it is clear how to generalize these paths to higher genus.

Fix a hyperbolic surface M and choose a simple closed geodesic $\emptyset$ on M. If we cut along this geodesic, twist left distance t and glue back along $\emptyset$, we define a new point, M_t, in T_g. This will be called the *time t twist deformation* of M; it is defined for any $\emptyset$, all $M \in T_g$ and all $t \in \boldsymbol{R}$. These paths are the generalization to higher genus of the horocycle paths in T_1. Since there are only a countable number of simple closed geodesics, the twists are not sufficient to fill out the whole of T_g. As in the genus 1 case, we need a more general deformation, determined by a geodesic foliation. Such foliations do not exist on hyperbolic surfaces, but partial foliations, called *geodesic laminations*, do.

DEFINITION. A *geodesic lamination* $\mathscr{L}$ on a hyperbolic surface M is a closed subset of M which is a union of simple geodesics. We assume that $\mathscr{L}$ can be covered by open sets U_i with maps $\emptyset_i$ to $\boldsymbol{R}^2$ sending $\mathscr{L} \cap U_i$ to horizontal arcs and with overlap maps preserving these horizontal arcs.

We further require that $\mathscr{L}$ possess a positive transverse Borel measure μ, finite on compact sets, whose support is all of $\mathscr{L}$. This implies that a local cross-section of $\mathscr{L}$ is either discrete or a Cantor set. We drop the distinction between the set and the transverse measure and denote both by μ. This measure allows us to define the intersection number, $i(\mu, \cdot)$, of μ with simple closed curves and arcs as in the case of foliations on the torus.

We denote the space of all such μ (allowing the trivial lamination $\mu \equiv 0$) by $\mathscr{ML}$; throwing out the zero lamination and factoring out by multiplication by positive scalars we have the space $\mathscr{PL}$ of projective classes of measured laminations. The simplest example of a geodesic lamination is a simple closed geodesic $\emptyset$ together with a constant times the counting measure. Then the intersection number with any closed geodesic γ is that constant times the number of intersections of $\emptyset$ and γ.

In the case of the torus any linear geodesic foliation can be approximated by ones with rational slope i.e. parallel copies of simple closed geodesics. The seminal result for higher genus is that this is still true.

THEOREM 2.1 (Thurston [22]). *$\mathscr{ML}$ is homeomorphic to $\boldsymbol{R}^{6g-6}$, $\mathscr{PL}$ to S^{6g-7}. Weighted simple closed curves are dense in $\mathscr{ML}$.*

This theorem allows us to generalize numerous concepts and operations from the set S of simple closed geodesics to general laminations. A deep observation of Thurston's is that the twist maps can be generalized to the case of a general $\mu \in \mathscr{ML}$. By definition, if $\mu = (\emptyset, c) \in S \times \boldsymbol{R}_+$, i.e., a weighted simple closed curve, then the time t deformation corresponding to μ is just the time tc twist deformation along the geodesic $\emptyset$ and is denoted by $\mathscr{E}_\mu(t)$.

PROPOSITION 2.2 ([10]). *Let $M \in T_g$ and $\mu \in \mathscr{ML}$ and suppose that $c_i\emptyset_i$ converges to μ in $\mathscr{ML}$. Then for all $t \leqslant T$, the curves $\mathscr{E}_{c_i\emptyset_i}(t)$, beginning at M, converge uniformly in T_g to a continuous curve $\mathscr{E}_\mu(t)$.*

The limit curve, $\mathscr{E}_\mu(t)$, is called the *time t earthquake deformation* of M determined by μ. On the surface itself, it is realized by shearing to the left along the leaves of μ by an amount proportional to the transverse measure. It is continuous away from the atoms of the measure (see [10] for details). This is the hyperbolic version of the shearing maps along irrational geodesic foliations discussed in Section I; the curves $\mathscr{E}_\mu(t)$ are the analogs of the general horocycle paths. Now there are enough curves to fill out T_g.

THEOREM 2.3 (Thurston, see [10] for proof). *There is a unique (left) earthquake path between any two points in T_g.*

This is the analog of the fact that, in $\boldsymbol{H}^2$, there is a unique (left) horocycle between any two points.

The key geometric fact about earthquake paths is that the geodesic length function $l_\emptyset$, $\emptyset \in S$, is convex along them.

THEOREM 2.4 ([10]). *The length function $l_\emptyset$: $T_g \to \boldsymbol{R}_+$ is convex along $\mathscr{E}_\mu(t)$ for any $\mu \in \mathscr{ML}$ and any $\emptyset \in S$. It is strictly convex iff $i(\mu, \emptyset) \neq 0$.*

Remark. Wolpert [26] computes the second derivative of $l_\emptyset$ along $\mathscr{E}_\mu(t)$ and shows that it is positive in the case $\mu \in S \times \boldsymbol{R}_+$ $i(\mu, \emptyset) \neq 0$. Together with limit arguments similar to those in [10] this result gives a second proof of Theorem 2.4.

Theorem 2.4 shows that for any finite collection $\{\emptyset_i\}$ of geodesics the sets $I_r = \{M \in T_g | \, l_{\emptyset_i}(M) \leqslant r, \forall i\}$ are convex. The $\{\emptyset_i\}$ are said to *fill up* M if all the components of $M - \bigcup \emptyset_i$ are disks. If the $\{\emptyset_i\}$ fill up M, it is easy to show that the sets $I_r \subset T_g$ are all compact. On the other hand, the lengths cannot all become arbitrarily small (since the curves intersect) so I_r is empty for sufficiently small r. Thus there is a smallest r for which I_r is non-empty; denote this by r_0.

We are now in a position to give a solution to the Nielsen realization problem analogous to that in the torus case.

Let G be any finite subgroup of Mod_g. Take any finite collection of simple closed curves, $\{\emptyset_i\}$, whose orbit under G, $\bigcup_{g \in G} \bigcup_i g\emptyset_i$, fills up M. Then for all r the set I_r is compact and G-invariant. Let r_0 be the smallest value such that I_{r_0} is non-empty. Choose a point in I_{r_0} with the property that the subset of curves (denoted $\{\tilde{\emptyset}_i\}$) whose length is exactly r_0 has the smallest cardinality. On any earthquake path which starts at this point and is contained in I_{r_0} either all the lengths of the subset of curves remain equal to r_0 or else they are strictly less than r_0 away from the endpoints (by convexity). The latter cannot happen by the minimality property of the subset so their lengths must be constant on all of I_{r_0}. It follows that the subset is also a union of G-orbits. There are two cases depending on whether or not the subset fills up M.

If the sub-collection $\{\tilde{\emptyset}_i\}$ does not fill up M, then there are simple closed geodesics, disjoint from the subset, dividing M into sub-surfaces that are permuted under G. In this case the proof follows by induction on the complexity of the surface. (The previous discussion is valid for the sub-surfaces since it holds generally for surfaces with geodesic boundary, elliptic, and parabolic points.) If the subset fills up M, then we claim that the corresponding set $\tilde{I}_{r_0}$, for the subset $\{\tilde{\emptyset}_i\}$ is a point. Since the $\{\tilde{\emptyset}_i\}$ are G-invariant $\tilde{I}_{r_0}$ is also G-invariant and we are done. To see that $\tilde{I}_{r_0}$ is a point note that any two points in it can be joined by an earthquake path $\mathscr{E}_\mu(t)$ in the set (by convexity) and $i(\mu, \emptyset_i) \neq 0$ for some i since the $\tilde{\emptyset}_i$ fill up M. By Theorem 2.4 $l_{\emptyset_i}$ is not constant and must be less than r_0 except at the endpoints of the path, a contradiction. Thus we have shown:

THEOREM 2.5 [10]. *Every finite subgroup of* Mod_g *acting on* T_g *has a fixed point.*

Remark. Tukia has recently announced a new proof of this theorem [23].

In fact, only the second case in the previous argument (where the sub-collection fills up M) can occur. This follows from a nice argument due to Maskit, Matelski, and Wolpert [15] (in a slightly different context). Given any simple closed geodesic $\emptyset$ on a hyperbolic surface M, one can cut along the geodesic and glue in an annulus, defining a new conformal structure. It is then possible to compare lengths in the hyperbolic metric conformally equivalent to the new structure with those in the original metric. The result is that any geodesic, not intersecting $\emptyset$ on the original surface, is shorter in the new hyperbolic metric.

Beginning with a collection of curves filling up M the subset whose lengths equal r_0 at a point in I_{r_0} must fill up M or else the lengths of the subset could all be decreased by the deformation above. The other lengths, already less than r_0, would remain so on an open set, contradicting the minimality of r_0. Thus we have:

COROLLARY 2.6. *Given a collection of simple closed geodesics, filling up M, the set I_{r_0} is always a point for the smallest value of r such that I_r is non-empty.*

COROLLARY 2.7. *Let $\{\emptyset_i\}$ be any collection of simple closed curves whose orbits under a finite group $G \subset \mathrm{Mod}_g$ minimally fill up M (i.e., no orbit can be deleted and the collection still fill up M). Then there is a fixed point in T_g under the action of G such that the lengths of the $\{\emptyset_i\}$ are all equal.*

If we recall that a sum of convex functions is convex, we see that the previous discussion applies equally well to sums of length functions. Let $l_{\underline{\emptyset}}$ denote the sum of $l_{\emptyset_i}$ where $\underline{\emptyset} = \{\emptyset_i\}$. If the $\{\emptyset_i\}$ fill up M, then $l_{\underline{\emptyset}}$ has a minimum in T_g. Since $i(\mu, \emptyset_i) \neq 0$, some i, for every $\mu \in \mathscr{ML}$, $l_{\underline{\emptyset}}$ is *strictly* convex along every $\mathscr{E}_\mu(t)$, so the minimum is unique.

THEOREM 2.8 [10]. *The sum of lengths function $l_{\underline{\emptyset}}$ has a unique minimum iff $\underline{\emptyset} = \{\emptyset_i\}$ fills up M.*

The "only if" part follows from the earlier argument of adding an annulus at a geodesic missing the $\{\emptyset_i\}$ if they do not fill up M. If $\{\emptyset_i\}$ is any G-invariant collection of geodesics, $G \subset \mathrm{Mod}_g$ finite, the unique minimum is a fixed point for G. This gives a more immediate proof of Theorem 2.5, the approach used in [9], [10]. Some information about the fixed points, i.e., Corollary 2.7, is lost in this case.

III

In this section we discuss some further geometric and analytic properties of the length function. One missing component in our comparison between hyperbolic space and T_g is the definition of a "geodesic" in terms of the length function. We describe reasonable candidates below; however, we have no suggestion for a metric on T_g in which they would be actually geodesics. There is some indication that they are at least closely connected to the Weil–Petersson metric, [12], [26].

First, we need to generalize the geodesic length function as in the torus case, from simple closed curves to measured geodesic laminations. This is defined as before to be the total mass of the product measure $dl \times d\mu$ where dl is the length measure along the leaves of μ, $\mu \in \mathscr{ML}$. The generalized length, l_μ, is as well-behaved as the lengths of simple closed geodesics.

THEOREM 3.1 [11], [12]. *The length function l_μ is convex along $\mathscr{E}_\nu(t)$, strictly convex iff $i(\nu, \mu) \neq 0$. It is a real analytic function on T_g, constant along the earthquake path $\mathscr{E}_\mu(t)$. For any fixed $M \in T_g$, l_μ is continuous as a function of $\mathscr{ML}$; in particular, $l_{c_i \emptyset_i} \to l_\mu$, uniformly on compact subsets of T_g, whenever $c_i \emptyset_i \to \mu$ in $\mathscr{ML}$, $c_i \emptyset_i \in S \times \boldsymbol{R}_+$.*

Given any two geodesic laminations μ and ν which together fill up M, we note that the average $(l_\mu + l_\nu)/2$ attains a minimum in T_g. By convexity the minimum is unique. In general, $tl_\mu + (1-t)l_\nu$ attains a unique minimum for any $t \in (0, 1)$. We call this one-parameter family of minima a *line of minima*. It is not hard to check that for T_1 they are precisely the geodesics in $\boldsymbol{H}^2$.

As $t \to 0$, $l_\nu \to 0$ and as $t \to 1$, $l_\mu \to 0$, so the lines are unbounded, and they are always transverse to the level sets of l_μ and l_ν. They are the traces of the one-parameter family of points of tangency of these level sets. Moreover, T_g is filled with these lines.

THEOREM 3.2 [12]. *Given any $\mu \in \mathscr{ML}$ and any point $M \in T_g$, there is a unique $\nu \in \mathscr{ML}$ such that $l_\mu + l_\nu$ has its unique minimum at M. In other words, there is a unique line of minima through M determined by μ.*

Hubbard and Masur ([5], see also [8]) proved the analogous result in the context of Teichmüller geodesics and measured foliations. In either case, the theorem can be interpreted as the generalization of the theorem in hyperbolic space of the existence of a unique line from any interior point to any given point at infinity. (There is the same measure-zero indeterminacy of the "sphere at infinity" here as in the case of the geometry of Teichmüller's metric.)

In the process of proving Theorem 3.2 we see that $\mathscr{ML}$ can be interpreted as being equal to the tangent or cotangent space at any point in T_g.

THEOREM 3.3 [12]. *The one-forms dl_μ, $\mu \in \mathscr{ML}$, are in 1–1 correspondence with the cotangent space of T_g at any point $M \in T_g$. The tangents to $\mathscr{E}_\mu(t)$ are similarly in 1–1 correspondence with the tangent space. For any two transverse μ, ν in $\mathscr{ML}$, filling up M, the condition $dl_\mu = -k\,dl_\nu$, $k \in (0, \infty)$ determines a unique line of minima in T_g.*

The last statement is the analog of the statement in $\boldsymbol{H}^2$ that geodesics are in 1–1 correspondence with pairs of points at infinity. (Note that the transversality and filling up conditions are necessary; however, for a set of μ's of full measure only the pairs (μ, ν) where $\nu = c\mu$, $c \in \boldsymbol{R}_+$, are ruled out by these conditions.)

The missing pieces in this program are the existence of a unique line of minima between any two points in T_g and some connection with a metric on T_g. There seems to be sufficient evidence for the following:

CONJECTURE. *There is a unique line of minima between any two points in T_g.*

In a series of lovely papers ([25], [26], [27], [28]) Wolpert has explored the relationship between the length functions of simple closed geodesics, twist maps, and the symplectic geometry of T_g determined by the (Kähler) Weil–Petersson metric. In particular he shows that the Kähler 2-form equals $\sum_{i=1}^{3g-3} dl_{\emptyset_i} \wedge d\theta_i$ where $\emptyset_i$ are $3g-3$ disjoint simple closed curves and θ_i are twist parameters at the $\emptyset_i$. (The co-ordinates $(l_{\emptyset_i}, \theta_i)$, $i = 1, 2, 3, \ldots, 3g-3$ are the classical Nielsen–Fenchel co-ordinates for T_g.) Furthermore, he shows that the vector field $\tau_\emptyset$ tangent to $\mathscr{E}_\emptyset(t)$, $\emptyset \in S$, is Hamiltonian and that the dual (with respect to the Weil–Petersson metric) to the one form $dl_\emptyset$, $\emptyset \in S$ is $i\tau_\emptyset$ ($i = \sqrt{-1}$). This close connection between the metric and the length function, together with the heuristic sense that the lines of minima should be orthogonal to the level sets of the length functions in some natural setting suggests the following:

Question. Is the gradient vector field of l_μ (in the Weil–Petersson metric) tangent to all the lines of minima determined by μ? (There is a unique such line through each point by Theorem 3.2.)

If the function $l_\mu + l_\nu$ attains its minimum at $M \in T_g$, then its Hessian determines a well-defined, positive-definite bilinear form B_μ. A positive answer to the question above is equivalent to the statement that the B_μ-orthogonal complement to the level set of l_μ (or l_ν) equals the orthogonal complement with respect to the Weil–Petersson metric. More generally, it would be very interesting to know how B_μ varies with μ and with the surface M. Information of this kind could well lead to a new metric on T_g and to further links between the geodesic length function and the geometry of Teichmüller space.

References

[1] Bers L., Finite Dimensional Teichmüller Spaces and Generalizations, *Bull. AMS* 5 (1981), pp. 131–172.

[2] Bers L., An Extremal Problem for Quasiconformal Mappings and a Theorem by Thurston, *Acta. Math.* **141** (1978), pp. 73–98.

[3] Bers L., Quasiconformal Mappings and Teichmüller's Theorem In: R. Nevan-

linna *et al.* (eds.), *Analytic Functions,* Princeton Univ. Press, Princeton, 1960, pp. 89–111.
[4] Bers L., The Action of the Modular Group on the Complex Boundary, Riemann Surfaces and Related Topics, *Ann. of Math. Studies* **97** (1981), pp. 33–52.
[5] Hubbard J. and Masur H., Quadratic Differentials and Measured Foliations, *Acta Math.* **142** (1979), pp. 221–274.
[6] Jorgensen T., Compact 3-Manifolds of Constant Negative Curvature Fibering over the Circle, *Ann. of Math.* **106** (1979), pp. 61–72.
[7] Jorgensen T., On Discrete Groups of Möbius Transformations, *Amer. J. of Math.* **98** (1976), pp. 739–749.
[8] Kerckhoff S., The Asymptotic Geometry of Teichmüller Space, *Topology* **19** (1980), pp. 23–41.
[9] Kerckhoff S., The Nielsen Realization Problem, *Bull. AMS* **2** (1980), pp. 452–454.
[10] Kerckoff S., The Nielsen Realization Problem, *Ann. of Math.* **117** (1983), pp. 235–265.
[11] Kerckhoff S., *Earthquakes Are Analytic,* preprint.
[12] Kerckhoff S., *Lines of Minima in Teichmüller Space,* preprint.
[13] Kerckhoff S., *Simplicial Systems for Interval Exchange Maps and Measured Foliations,* preprint.
[14] Kravetz S., On the Geometry of Teichmüller Space and the Structure of Their Modular Groups, *Ann. Acad. Sci. Fenn.* **278** (1959), pp. 1–35.
[15] Maskit B., Matelski P., and Wolpert S., *Minimal Geodesics on Riemann Surfaces,* preprint.
[16] Masur H., On a Class of Geodesics in Teichmüller Space, *Ann. of Math.* **102** (1975), pp. 205–221.
[17] Masur H., Uniquely Ergodic Quadratic Differentials, *Comm. Math. Helv.* **55** (1980), pp. 255–266.
[18] Masur H., Interval Exchange Maps and Measured Foliations, *An. of Math.* **115** (1982), pp. 169–200.
[19] Masur H., Two Boundaries of Teichmüller Space. *Duke Math. J.* **49** (1982), pp. 183–190.
[20] Rees M., An Alternative Approach to the Ergodic Theory of Measured Foliations on Surfaces, *Ergodic Theory and Dynamical Systems* **1** (1981), pp. 461–488.
[21] Thurston W., *On the Geometry and Dynamics of Diffeomorphisms of Surfaces, I,* preprint.
[22] Thurston W., *Three-Dimensional Topology and Geometry,* Bound notes, Princeton Univ.
[23] Tukia P., *Homeomorphic Conjugates of Fuchsian Groups,* preprint.
[24] Veech W., Gauss Measures for Transformations on the Space of Interval Exchange Maps, *Ann. of Math.* **115** (1982), pp. 201–242.
[25] Wolpert S., The Fenchel–Nielsen Deformation, *Ann. of Math.* **115** (1982), pp. 501–528.
[26] Wolpert S., On the Symplectic Geometry of Deformations of a Hyperbolic Surface, *Ann. of Math.* **117** (1983), pp. 207–234.
[27] Wolpert S., *On the Kähler Form of the Moduli Space of once Punctured Tori,* preprint.
[28] Wolpert S., *On the Weil-Petersson Geometry of the Moduli Space of Curves,* preprint.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305, U.S.A.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

WEN-HSIUNG LIN *

Some Remarks on the Kervaire Invariant Conjecture

Let A denote the mod 2 Steenrod algebra. Let $\psi_{i,i}$ be the secondary mod 2 cohomology operation based on the Adem relation

$$\mathrm{Sq}^{2^i}\mathrm{Sq}^{2^i}+\sum_{j=0}^{i-1}\mathrm{Sq}^{2^{i+1}-2^j}\mathrm{Sq}^{2^j}=0$$

in A as described by Adams in [2]. Let S^n denote the sphere spectrum in stable dimension n ([3]). Call a homotopy class $S^{2^{i+1}-2}\to S^0$ θ_i if $\psi_{i,i}$ is non-zero in $H^*(S^0\cup_{\theta_i}e^{2^{i+1}-1})$ where $H^*(\ )$ is the mod 2 cohomology functor. θ_i exists if and only if in the mod 2 Adams spectral sequence $\{E_r^{s,t}\}$ for the stable homotopy groups of spheres ([1]) the class $h_i^2 \in \mathrm{Ext}_A^{2,2^{i+1}}(H^*(S^0),\boldsymbol{Z}_2)=E_2^{2,2^{i+1}}$ survives the spectral sequence. It is classical ([4]) that θ_i exists for $0\leqslant i\leqslant 3$: h_0^2 detects 4ι where ι generates $\pi_0(S^0)=\boldsymbol{Z}$ while h_1^2, h_2^2 and h_3^2 detect the products η^2, ν^2 and σ^2 of the Hopf classes η, ν and σ, respectively. Mahowald and Tangora ([8]) have shown that θ_4 also exists. The Kervaire invariant conjecture, in its homotopy version, asserts that θ_i also exists for $i\geqslant 5$. A good exposition of the conjecture, including a description of the geometric roots, can be found in the recent paper by M. G. Barratt, J. D. S. Jones and M. E. Mahowald [5]; some of the results of two of these authors announced in [9], which were obtained in an effort to settle the problem, are proved in [5].

From the homotopy point of view it is natural to study a stronger version of the problem: Does there exist a θ_i with $2\theta_i=0$? This stronger version was put forth by Barratt and Mahowald in [9]. Here we give two evidences for the truth of this stronger conjecture. Both involve the differentials in the Adams spectral sequence.

* Supported by Academia Sinica, China-Taiwan.

The first one of these describes non-trivial cohomology extensions in certain 2-stage Postnikov systems. Let $i \geqslant 4$. Consider the 2-stage Postnikov system

$$\begin{array}{ll} \Sigma^{-1}L \xrightarrow{j} E & \\ \quad\quad\ \downarrow p & \\ \quad\quad\ K \xrightarrow{f=(\mathrm{Sq}^2,\mathrm{Sq}^2\mathrm{Sq}^1,\mathrm{Sq}^4,\ldots,\mathrm{Sq}^{2^i})} L \end{array} \tag{1}$$

where K denotes the Eilenberg–Maclane spectrum $K(\mathbf{Z}_2)$ and $L = \Sigma^2K \times \times\Sigma^3K \times \Sigma^4K \times \ldots \times \Sigma^{2^i}K$. The group $H^*(K)/\mathrm{im} f^*$, which is embedded in $H^*(E)$, is generated by $1, \mathrm{Sq}^1, \mathrm{Sq}^{2^{i+1}}$ and $\mathrm{Sq}^{2^{i+1}+1}$ for $* \leqslant 2^{i+1}+1$. It is easy to see that any minimal set of generators of $H^*(E)$ over A for $3 \leqslant * \leqslant 2^{i+1}-1$ is one-to-one correspondent with a $\mathbf{Z}_2$-base of $\mathrm{Ext}_A^{2,*+1}\big(H^*(S^0\cup_{2\iota}e^1), \mathbf{Z}_2\big)$ where $S^0\cup_{2\iota}e^1$ is the mapping cone of $2\iota : S^0 \to S^0$. We fix any such set. Let $h_k \in \mathrm{Ext}_A^{1,2^k}\big(H^*(S^0), \mathbf{Z}_2\big)$ be the classes corresponding to the generators $\mathrm{Sq}^{2^k} \in A$; their images in $\mathrm{Ext}_A^{1,*}\big(H^*(S^0\cup_{2\iota}e^1), \mathbf{Z}_2\big)$ are also denoted by h_k.

Adams ([2]) has shown that $h_i^2h_1 \neq 0$ in $\mathrm{Ext}_A^{3,2^{i+1}+2}\big(H^*(S^0\cup_{2\iota}e^1), \mathbf{Z}_2\big)$ (since $i \geqslant 4$). This implies

$$j^*(\mathrm{Sq}^2\varphi_{i,i} + \mathrm{Sq}^{2^i}\varphi_{i,1} + \ldots) = 0$$

where $\varphi_{m,l}$ are the classes in $H^*(E)$ corresponding to the Adams invariants $h_mh_l \in \mathrm{Ext}_A^{2,*}\big(H^*(S^0\cup_{2\iota}e^1), \mathbf{Z}_2\big)$ (in dimensions $\leqslant 2^{i+1}-1$). So

$$\mathrm{Sq}^2\varphi_{i,i} + \mathrm{Sq}^{2^i}\varphi_{i,1} + \ldots = \lambda\mathrm{Sq}^{2^{i+1}+1} \tag{2}$$

for some coefficient λ which is 0 or 1.

THEOREM A. $\lambda = 1$ *for all* $i \geqslant 4$.

These non-trivial extensions are connected with the strong Kervaire invariant conjecture in the following way. The stronger conjecture is equivalent to asserting that $h_{i+1} \in \mathrm{Ext}_A^{1,2^{i+1}}\big(H^*(S^0\cup_{2\iota}e^1), \mathbf{Z}_2\big)$ detects homotopy elements in $\pi_{2^{i+1}-1}(S^0\cup_{2\iota}e^1)$ for all $i \geqslant 4$ (see [9]). If there is a homotopy class $\{h_{i+1}\}$: $S^{2^{i+1}-1} \to S^0\cup_{2\iota}e^1$ detected by h_{i+1}, let $X = S^0\cup_{2\iota}e^1\cup e^{2^{i+1}}$ be its mapping cone. The Steenrod operation $\mathrm{Sq}^{2^{i-1}}$ is non-zero in $H^*(X)$; using this fact, one verifies that there is only one non-zero class in $\mathrm{Ext}_A^{1,2^{i+1}+1}\big(H^*(X), \mathbf{Z}_2\big)$ and that this class corresponds to $\mathrm{Sq}^{2^{i+1}+1} \in H^*(E)$ in the system (1); we denote it by γ. The composite

$$S^{2^{i+1}-1} \xrightarrow{\{h_{i+1}\}} S^0\cup_{2\iota}e^1 \to S^1$$

is a $\theta_i \in \pi_{2^{i+1}-1}(S^1)$ with $2\theta_i = 0$, as shown by Barratt and Mahowald in [9]. Let $\bar\theta_i$ be the image in $\pi_{2^{i+1}-1}(S^0\cup_{2\iota}e^1)$ of the desuspension of θ_i under the inclusion $S^0 \to S^0\cup_{2\iota}e^1$. Then, since $\eta\theta_i$ is contained in the Toda bracket $\langle 2\iota, \theta_i, 2\iota\rangle$ ([11], III, 3.7), $\eta\bar{\theta}_i = 2\{h_{i+1}\}$, which is detected by $h_i^2 h_1 \in \mathrm{Ext}_A^{3,2^{i+1}+2}(H^*(S^0 \cup_{2\iota} e^1), \boldsymbol{Z}_2)$. The image $\overline{h_i^2 h_1}$ of $h_i^2 h_1$ in $\mathrm{Ext}_A^{3,2^{i+1}+2}(H^*(X), \boldsymbol{Z}_2)$ is non-zero and is a permanent cycle in the Adams spectral sequence for X (since $h_i^2 h_1$ is so for $S^0 \cup_{2\iota} e^1$). If this class is non-zero in E_∞, then the image of $\eta\bar{\theta}_i$ in $\pi_{2^{i+1}-1}(X)$ projects to it, which is impossible since $\eta\bar{\theta}_i = 2\{h_{i+1}\}$. Thus $\overline{h_i^2 h_1}$ has to be a boundary and the only possibility is that $d_2(\gamma) = \overline{h_i^2 h_1}$, which means exactly that the extension (2) is non-trivial.

Theorem A is proved in [6] without the assumption on h_{i+1}. The significance of the result is that "the differentials $d_2(\gamma) = \overline{h_i^2 h_1}$ exist" although we do not know anything about $\{h_{i+1}\}$.

To describe the second evidence, let i, j be two integers such that $i \geqslant 5$, $j \geqslant 5$ and $h_i^2 h_j^2 h_1^2 \neq 0$ in $\mathrm{Ext}_A^{6,2^{i+1}+2^{j+1}+2}(H^*(S^0), \boldsymbol{Z}_2)$. It is shown in [7] that $h_i^2 h_j^2 h_1^2 \neq 0$ if i and j are far apart. Suppose θ_i and θ_j with $2\theta_i = 2\theta_j = 0$ exist. Choose an element η_{i+1} in the Toda bracket $\langle \eta, 2\iota, \theta_i\rangle$. Then η_{i+1} is detected by $h_{i+1}h_1 \in \mathrm{Ext}_A^{2,2^{i+1}+2}(H^*(S^0), \boldsymbol{Z}_2)$. We have $2\eta_{i+1} = 2\langle \eta, 2\iota, \theta_i\rangle = \langle 2\iota, \eta, 2\iota\rangle\theta_i = \eta^2\theta_i$, which is detected by $h_i^2 h_1^2$. Thus $2\eta_{i+1}\theta_j$ is detected by $h_i^2 h_j^2 h_1^2$. But $2\theta_j = 0$; so $2\eta_{i+1}\theta_j = 0$. This implies that $h_i^2 h_j^2 h_1^2$ is a boundary in the Adams spectral sequence for S^0. If we perform formal substitutions "$h_{j+1} = 2\theta_j$" and "$h_{i+1}h_1 = \eta_{i+1}$", we find that it must be $h_{i+1}h_{j+1}h_1$ that hits $h_i^2 h_j^2 h_1^2$; i.e., $d_3(h_{i+1}h_{j+1}h_1) = h_i^2 h_j^2 h_1^2$. Our result is that this is indeed the case.

THEOREM B. *$d_3(h_{i+1}h_{j+1}h_1) = h_i^2 h_j^2 h_1^2$ in the Adams spectral sequence for S^0.*

Here the integers i and $j \geqslant 5$ and are far apart. We assume $i \gg j$.

We indicate a proof of B. Consider a minimal (mod 2) Adams resolution of S^0 up to the second stage

$$\begin{array}{ccccc}
\Sigma^{-1}K_1 & \xrightarrow{j_2} & X_2 & & \\
 & & \downarrow{\scriptstyle p_2} & & \\
\Sigma^{-1}K_0 & \xrightarrow{j_1} & X_1 & \xrightarrow{f_1} & K_1 \\
 & & \downarrow{\scriptstyle p_1} & & \\
 & & S^0 & \xrightarrow{f_0} & K_0 = K
\end{array}$$

where $f_1 = (\mathrm{Sq}^1, \mathrm{Sq}^2, \mathrm{Sq}^4, \ldots, \mathrm{Sq}^{2^i}, \ldots)$ and $K_1 = K \times \Sigma K \times \Sigma^3 K \times \ldots \ldots \times \Sigma^{2^i-1}K \times \ldots$ Then $\mathrm{Ext}_A^{s,t}(H^*(X_2), \boldsymbol{Z}_2) \simeq \mathrm{Ext}_A^{s+2,t+2}(H^*(S^0), \boldsymbol{Z}_2)$ and the

Adams spectral sequence $\{E_r^{s,t}(X_2)\}$ for X_2 is a faithful portion of the Adams spectral sequence for S^0. In particular, the differential $d_3(h_{i+1}h_{j+1}h_1) = h_i^2h_j^2h_1^2$ can be considered as in $\{E_r^{*,*}(X_2)\}$.

To prove Theorem B we will raise the Adams filtration of $h_{i+1}h_1 \in \mathrm{Ext}_A^{0,*}(H^*(X_2), \boldsymbol{Z}_2)$ by 1 by considering a spectrum X_2' which is equal to X_2 in a certain range of dimensions so that B is equivalent to $d_2(h_{j+1}h_{i+1}h_1) = h_i^2h_j^2h_1^2$ in $\{E_r^{*,*}(X_2')\}$.

X_2' is obtained by killing the A-module generators of $H^*(X_1) = \Sigma^{-1}\bar{A}$ ($\bar{A}$ = augmentation ideal of A) only up to Sq^{2^i}. More precisely, X_2' is the fiber of $f_1' = (\mathrm{Sq}^1, \mathrm{Sq}^2, \ldots, \mathrm{Sq}^{2^i})$: $X_1 \to K_1' = K \times \Sigma K \times \Sigma^3 K \times \ldots \ldots \times \Sigma^{2^i-1}K$ so that there is a cofibering $X_2' \xrightarrow{p_2'} X_1 \xrightarrow{f_1'} K_1'$. The group $H^*(X_1)/\mathrm{im} f_1'^*$, which is embedded in $H^*(X_2')$ via $(p_2')^*$, is generated by $\mathrm{Sq}^{2^{i+1}}$ for $* \leqslant 2^{i+1}-1$. We have $H^*(X_2') \simeq H^*(X_2)$ for $* \leqslant 2^{i+1}-2$; so $B = \{h_l h_m \mid 0 \leqslant l \leqslant m \leqslant i,\ m \neq l+1\}$ is a $\boldsymbol{Z}_2$-base for $\mathrm{Ext}_A^{0,*}(H^*(X_2'), \boldsymbol{Z}_2)$ in this range of dimensions. B can be considered as a subset of $H^*(X_2')$ and forms a minimal set of generators of $H^*(X_2')$ over A for $* \leqslant 2^{i+1}-2$. Adams ([2]) has shown that there is a non-trivial extension

$$\mathrm{Sq}^1(h_i^2) + \mathrm{Sq}^{2^i}(h_ih_0) + \ldots = \mathrm{Sq}^{2^{i+1}} \tag{$*$}$$

in $H^*(X_2')$. Using this, one verifies that there is only one non-zero class in $\mathrm{Ext}_A^{1,2^{i+1}+1}(H^*(X_2'), \boldsymbol{Z}_2)$ and that this class corresponds to the class $h_{i+1}h_1$ in $\mathrm{Ext}_A^{0,2^{i+1}}(H^*(X_2), \boldsymbol{Z}_2)$; we denote it by $\overline{h_{i+1}h_1}$. One also verifies that $h_i^2h_1^2 \neq 0$ in $\mathrm{Ext}_A^{2,2^{i+1}+2}(H^*(X_2'), \boldsymbol{Z}_2)$ (since $i \geqslant 5$) and that $h_0\overline{h_{i+1}h_1} = h_i^2h_1^2$. (Perhaps, the easiest way to see these facts is to observe that $\overline{h_{i+1}h_1}$ is contained in the Massey product $\langle h_1, h_0, h_i^2\rangle$ formed in $E_3^{*,*}(X_2)$ and that $h_0\langle h_1, h_0, h_i^2\rangle = \langle h_0, h_1, h_0\rangle h_i^2 = h_1^2h_i^2$; these correspond to $\eta_{i+1} \in \langle \eta, 2\iota, \theta_i\rangle$ and $2\langle \eta, 2\iota, \theta_i\rangle = \langle 2\iota, \eta, 2\iota\rangle\theta_i = \eta^2\theta_i$ in the homotopy level.) Since $i \gg j$, it is not difficult to show that $h_{j+1}\overline{h_{i+1}h_1} \neq 0$ in $\mathrm{Ext}_A^{2,*}(H^*(X_2'), \boldsymbol{Z}_2)$ and that $h_j^2h_i^2h_1^2 \neq 0$ in $\mathrm{Ext}_A^{4,*}(H^*(X_2'), \boldsymbol{Z}_2)$. The differential $d_2(h_{j+1}\overline{h_{i+1}h_1}) = h_j^2h_i^2h_1^2$ now follows from the Adams differential $d_2(h_{j+1}) = h_j^2h_0$ ($j \geqslant 5$), $h_0\overline{h_{i+1}h_1} = h_i^2h_1^2$ and the fact that $\overline{h_{i+1}h_1}$ survives the Adams spectral sequence for X_2', as proved by Mahowald in [10]. This proves B.

It is also shown in [7] that $h_{i_1}^2h_{i_2}^2\ldots h_{i_n}^2 \neq 0$ in $\mathrm{Ext}_A^{*,*}(H^*(S^0), \boldsymbol{Z}_2)$ for any increasing sequence $\{i_1, i_2, \ldots, i_n\}$ of positive integers such that i_j are far apart from one another. There is no difficulty in verifying that $d_3(h_i^2) = 0$ for all i ($d_2(h_i^2) = 0$ follows from the multiplicative property of the Adams spectral sequence for S^0). Then the result B shows that for $i_1 = 1$ all the classes $h_1^2 h_{i_2}^2 \dots h_{i_n}^2$ are boundaries in the Adams spectral sequence $\{E_r^{*,*}(S^0)\}$ $(n \geqslant 3)$. If the Kervaire invariant conjecture is true then the classes $h_{i_1}^2 h_{i_2}^2 \dots h_{i_n}^2$ are permanent cycles in $\{E_r^{*,*}(S^0)\}$ and if $i_1 > 1$, there seems no reason for these cycles to be boundaries.

References

[1] Adams J. F., On the structure and applications of the Steenrod algebra, *Comm. Math. Helv.* **32** (1958), pp. 180–214.

[2] Adams J. F., On the non-existence of elements of Hopf invariant one, *Ann. Math.* **72** (1960), pp. 20–104.

[3] Adams J. F., *Stable homotopy and generalized homology*, University of Chicago Press, 1974.

[4] Adem J., The relations on Steenrod powers of cohomology classes, in: *Algebraic Geometry and Topology*, Princeton, 1957, pp. 191–238.

[5] Barratt M. G., Jones J. D. S., and Mahowald M. E., The Kervaire invariant problem, to appear in the *Proceedings of a conference at Northwestern University*.

[6] Lin W. H., Cohomology extensions in certain 2-stage Postnikov systems, to appear in the *Proceedings of a conference at Northwestern University*.

[7] Lin W. H., Cohomology of the Steenrod algebra, to appear in the *Journal of Pure and Applied Algebra*.

[8] Mahowald M. E. and Tangora M. C., Some differentials in the Adams spectral sequence, *Topology* **6** (1967), pp. 349–369.

[9] Mahowald M. E., Some remarks on the Kervaire invariant problem from the homotopy point of view, *Proc. Symp. Pure Math. AMS* **22** (1971), pp. 165–169.

[10] Mahowald M. E., A new infinite family in ${}_2\pi_*^S$, *Topology* **16** (1977), pp. 249–256.

[11] Toda H., *Composition methods in homotopy groups of spheres*, Ann. Math. Studies No. **49**, Princeton University Press, Princeton, 1962.

DEPARTMENT OF MATHEMATICS
NATIONAL TSING HUA UNIVERSITY
HSINCHU, TAIWAN 300

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

JULIUS L. SHANESON

Linear Algebra, Topology and Number Theory

In 1935, G. de Rham explicitly proposed the following conjecture:

Let A and B be orthogonal transformations of R^n; i.e. $A, B \in O(n; R)$, and suppose that A and B are topologically similar. Then A and B will be linearly similar.

That A and B are topologically similar (write $A \sim_t B$) means that there is a homeomorphism $\varphi: R^n \to R^n$ so that

$$\varphi A \varphi^{-1} = B; \tag{1}$$

i.e. A and B are conjugate in $\mathrm{Homeo}(R^n)$. Of course, linear similarity (write $A \sim B$) means that A and B are conjugate in $\mathrm{Gl}(n; R)$, i.e. φ can be taken to be linear.

By 1960 P. A. Smith had raised a fundamental question on *smooth* transformation groups. The way Smith mentioned this question in print (in a footnote to a discussion on fixed point free dihedral group actions on S^4) has left some room for disagreement on what his precise statement might have been. Here is a version of Smith's question, stated as a conjecture, which, on *purely mathematical grounds* (to be elaborated at several points below), seems the most natural to me:

Let G be a group that acts smoothly on the (homotopy) sphere Σ. Assume that for $H \subset G$, the fixed set Σ^H of H is either discrete or connected. Let $\Sigma^G = \{x, y\}$. Then the local linear representations of G on the tangent spaces Σ_x and Σ_y are linearly similar.

(The local representations of G at x and y are obtained by differentiating the action of G at x and y.)

P. A. Smith proved that for H a p-group, p a prime, acting on a mod p homology sphere Σ, the fixed point set Σ^H would also be a mod p homology sphere. Results of this type are generically called "Smith theory".

Thus Smith theory implies that for G a p-group, the hypothesis on Σ^H, $H \subset G$, is always satisfied. We shall say that an action of any group G on Σ with Σ^H discrete or connected for all $H \subset G$ is of *Smith type*. We shall also say that the representations at fixed points of a smooth action of Smith type on a homotopy sphere are *Smith* equivalent. Then the above conjecture says that Smith equivalence and linear similarity are the same.

This paper will discuss the work of many authors on these conjectures and some related questions. In particular, Cappell and the present author found counterexamples[1] and also obtained classification results on topological similarity and on Smith equivalence, at least for certain classes of representations. It seems to me that topological similarity is a much more fundamental relationship than Smith equivalence. Still there is some (partly conjectural) connection between Smith equivalence and topological similarity, and the classifications up to Smith equivalence and topological similarity are closely related to some phenomena in number theory.

§ 1. Some special cases of de Rham's conjecture

De Rham himself observed that $A \sim_t B$ implies that A and B have the same non-periodic eigenvalues, with the same multiplicities. Hence it always suffices to consider *periodic* orthogonal transformations or even, more generally, orthogonal representation of arbitrary (not just cyclic) groups. De Rham also proved the following:

1.1. THEOREM. *Suppose in* (1) *that* $\varphi(S^{n-1}) = S^{n-1}$, S^{n-1} *the unit sphere of* R^n, *and suppose that* $\varphi|S^{n-1}$ *and its inverse are smooth.*[2] *Then* A *and* B *are linearly similar.*

Note that if the theorem were valid without the smoothness condition, de Rham's conjecture would follow.

De Rham's argument is easiest to understand when A and B are periodic and induce free actions of a cyclic group on S^{n-1}. The quotient spaces L_A and L_B will then be "lens spaces", and the existence of a φ in (1) that is smooth on S^{n-1} implies that L_A and L_B are diffeomorphic.

---

[1] Consider the statement obtained from the above conjecture by deleting the requirement on Σ^H. Counterexamples to this version of Smith's question were first discovered by Petrie, with G a group of odd order having at least 3 non-cyclic Sylow subgroups.

[2] Or even just P.L. [30].

But de Rham showed that diffeomorphic lens spaces are linearly isometric; he did this by showing that a lens space is determined up to linear isometry by its Reidemeister torsion, a diffeomorphism invariant. This also uses a highly non-trivial bit of number theory, the Franz independence lemma. (See [26] for an account of the classification of lens spaces.) It follows that $A \sim B$. Actually this argument works, for the free case, even if φ is only a homeomorphism, as from [23] it follows that Reidemeister torsion is a topological invariant.

If A and B are not free, the quotient spaces L_A and L_B will be stratified spaces. But a *smooth* φ will induce a homeomorphism that not only preserves the strata but also behaves well on neighborhoods of the strata. This allows removal of the interior of a neighborhood of one strata in the next to obtain from φ a diffeomorphism of *closed* complements of strata. For these closed complements (which will be disk bundles over lens spaces), Reidemeister torsion will be defined. The above argument can then be applied inductively.

Kuiper and Robbin obtained further results. The main point of their work was to show how de Rham's conjecture would imply a complete classification of linear maps, up to topological similarity. But, using arguments of an elementary nature, they also proved the conjecture when the periodic eigenvalues have period at most six.

Schultz proved de Rham's conjecture for A and B of period p^a or $2p^a$, p an odd prime. To do this, he used Sullivan's KO-orientation, away from the prime two, for topological Euclidean space bundles. Sullivan's result depends upon the smooth or piecewise linear transversality principle. From the topological invariance of the KO-orientation, Schultz concludes that the forgetful map

$$a_G\colon \mathrm{KO}(B_G) \to K_{\mathrm{TOP}}(B_G)$$

is a monomorphism, for G a group of *odd* order. But according to [1], [4], $\mathrm{KO}(B_G)$ is the completion $\mathrm{RO}\hat{\ }(G)$ of the real representation ring of $\mathrm{RO}(G)$ with respect to the ideal of elements of virtual dimension zero. It follows that topologically similar representations represent the same element in $\mathrm{RO}\hat{\ }(G)$. Schultz then completes the argument by showing that for G cyclic of order p^a, $\mathrm{RO}(G) \to \mathrm{RO}\hat{\ }(G)$ is a monomorphism. This does not hold for G not a p-group.

The next major development, the results of Cappell and myself, will be discussed in the next section. But the following (unpublished) result of ours should be mentioned here: *If A, $B \in O(n)$ are topologically similar* *and* $n \leqslant 5$, *then* A *and* B *are linearly similar.* Lately we have begun to believe that we can prove the same thing through dimension eight.

Subsequently, Madsen–Rothenberg and Hsiang–Pardon proved theorems that include the following important result:[3] *Topological similarity implies linear similarity for rotations of odd order.* Hsiang and Pardon apply the most modern techniques in geometric topology, with the (moral) aim of "smoothing" a topological similarity, so that 1.1 would then hold. A further description of their work will be given in the section on Smith equivalence.

The main point of the work of Madsen–Rothenberg is an equivariant version of Sullivan's result; they show that locally linear topological G-vector bundles are oriented over the homology theory $\mathrm{KG}_G(\cdot)\otimes Z[1/2]$. To do so, they need to derive a stable transversality principle. They treat the piecewise linear case first and then they derive the result for the topological case by consideration of the "obstructions" to making everything piecewise linear. It should be noted that they use, among other things, the results of [2].

§ 2. Non-linear similarity

Cappell and I first obtained $A, B \in O(n)$, $A \sim_t B$ but A and B not linearly similar.

2.1. Theorem. *For each* $q > 1$, *there exist* $A_q, B_q \in O(9)$; *periodic of period* $4q$, *with* $A_q \sim_t B_q$ *but* A_q *and* B_q *not linearly similar.*

Actually we obtained complete neccessary and sufficient conditions for topological similarity for a certain special but interesting class of matrices. Let us say that $A \in O(n)$ of period $4q$ is *pseudo-free* if it has $+1$ or -1 (but not both) as an eigenvalue of multiplicity 1, and if all other eigenvalues are primitive $4q$th roots of 1. If A is *pseudo-free* and $B \sim_t A$, it can be shown that B is pseudo-free also.

Let T be a fixed multiplicative generator of Z_{4q}, and $Z[Z_{4q}] = Z[T\,|\,T^{4q} = 1]$ be the integral group ring. Let $R_q = Z[Z_{4q}]/(1+T+\ldots+T^{4q-1})$ be the quotient by the indicated ideal, and let $\gamma\colon Z[Z_{4q}]\to R_q$ be the quotient map. Let $\exp(2\pi i a_j/4q)$, $1 \leqslant j \leqslant m$, be the eigenvalues of $A \in O(2m+1)$ with $a_i \equiv 1+2q[a_i/2q] \pmod 4$, and let

$$\Delta_A = \gamma\Big(\prod_{j=1}^{m} (T^{r_j}-1)\Big),$$

[3] H.–P.'s results also apply when the 2-primary parts of the rotations satisfy certain special conditions.

where $a_j r_j \equiv 1 \pmod{4q}$. Note that the quotient Δ_A/Δ_B makes sense in R_q and is a ("cyclotomic") unit. Here is one form of the classification result of Cappell and myself:

2.2. THEOREM. *Let A and B be pseudo-free elements of $O(2m+1)$ of period $4q$. Then $A \sim_t B$ if and only if -1 is an eigenvalue of A and B and there is a unit $u(T) \in Z[Z_{4q}]^\times$ so that*

$$\Delta_A/\Delta_B = \gamma\big(T^\varepsilon u(T)u(-T)^{-1}\big), \qquad \varepsilon = 0 \text{ or } 2q.$$

As an example, let $R(\theta)$ be rotation of the plane thru an angle of $0 < \theta < \pi$ radians, and let

$$M(\theta_1, \theta_2, \ldots, \theta_k) = \begin{bmatrix} R(\theta_1) & & & & & & & & & \\ & R(\theta_1) & & & & & & & 0 & \\ & & R(\theta_2) & & & & & & & \\ & & & R(\theta_2) & & & & & & \\ & & & & \cdot & & & & & \\ & & & & & \cdot & & & & \\ & & & & & & \cdot & & & \\ & & & & & & & R(\theta_k) & & \\ & 0 & & & & & & & R(\theta_k) & \\ & & & & & & & & & -1 \end{bmatrix}.$$

2.2.1. COROLLARY. *Suppose $\theta_i/2\pi = a_i/4q$, $(a_i, 4q) = 1$. Then*

$$M(\pi-\theta_1, \pi-\theta_2, \ldots, \pi-\theta_s, \theta_{s+1}, \ldots, \theta_k) \sim_t M(\theta_1, \theta_2, \ldots, \theta_k),$$

provided that s is even.

Proof. We have to consider the product (in R_q) of cyclotomic units,

$$\prod_{i=1}^{s} [(T^{2q+r_i}-1)/(T^{r_i}-1)]^2.$$

But $(T^r+1)/(T^{2q+r}+1) = [(T^{2q+r}-1)/(T^r-1)]$ in R_q. Hence, if $v_i(T) = (T^{2q+r_i}-1)/(T^{r_i}-1)$, then the above product has the form

$$\prod_{i=1}^{s} v_i(T)v_i(-T)^{-1}.$$

There is a fiber product diagram

$$\begin{array}{ccc} Z[Z_{4q}] & \xrightarrow{\gamma} & R_q \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle \bar{a}} \\ Z & \longrightarrow & Z/4qZ, \end{array}$$

a the augmention, and $\bar{a}(v_i(T)) = 2q+1$. Since $(2q+1)^s \equiv 1 \pmod{4q}$ for s even, it follows that there is $u(T) \in Z[Z_{4q}]^\times$ with $\gamma(u(T)) = \prod_1^s v_i(T)$.

In the case q odd, the algebraic condition in 2.1 can be reduced to some purely elementary mod two number-theoretic equations. I will discuss this more in § 4. Let me just remark at this point that one can apply some number-theoretic considerations to show that for q a power of 2 or for $q < 28$ (but *not* for $q = 29$), the *only* topologically similar pseudo-free orthogonal transformations are the ones given in the Corollary.

To quickly describe the idea of the work of Cappell and myself, let

$$A = \begin{bmatrix} A_0 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} B_0 & 0 \\ 0 & -1 \end{bmatrix}.$$

Now, A_0 and B_0 will be free elements of $O(2m)$, A and B as in 2.1. Suppose for the sake of illustration that $A_0 \sim_t B_0$, and let φ be a homeomorphism with $\varphi A_0 \varphi^{-1} = B_0$. Let L_{A_0}, L_{B_0} be the quotients of S^{2m-1} by A_0 and B_0, respectively; those are *lens space*. The quotient of the region between S^{2m-1} and the image under φ of a very large sphere in R^{2m} will be a *topological* h-cobordism of L_{A_0} and L_{B_0}. By a theorem of Atiyah–Bott, applied to the topological case using the Kirby–Siebenmann theory (see [26] and compare [35]), this implies $L_{A_0} \cong L_{B_0}$ and so A_0 and B_0 would have to be linearly similar. Thus de Rham's conjecture actually holds for *free* elements of $O(2m)$.

A little contemplation with the above in mind shows that to study topological similarity for A and B, the right object to consider is a h-cobordism between unoriented interval bundles over L_{A_0} and L_{B_0}. Actually, one has to worry a little about the difference between open and closed interval bundles.[4] The algebraic condition in 2.1 is actually necessary and sufficient for the closed interval bundles to be h-cobordant. The proof makes use of generalized Browder–Livesay groups, which are the obstruction groups to a type of codimension-one surgery, as well as some calculations of normal invariants of lens spaces.[5]

Clearly, the results discussed so far leave many cases open. Stably, our understanding is much better. Two group representations

$$\varrho_i \colon G \to O(n)$$

[4] This could have been dealt with using material from [20], had it been available at the time of our work.

[5] In fact, the algebraic material referred to above and in §4 also leads to computations of various L^h-surgery groups.

will be called *topologically similar* if there exists a homeomorphism φ with $\varphi\varrho_1(g)\varphi^{-1} = \varrho_2(g)$ for all $g \in G$. The above results have obvious translations to the topological classification of pseudo-free representations of cyclic groups. Further, topological similarity is preserved by the following geometric operations: stabilization (replace ϱ_i by $\varrho_i + \theta$), induction from a subgroup, restriction, and composition with a homomorphism. One could ask if all pairs of topologically similar representations can be generated from the pseudo-free case by those operations. Various unstable phenomena make this apparently untrue (e.g. [15]). But it may be true for stable topological equivalence.

Stably and "modulo 2-torsion", we have complete determination of the topological classification of group representations, and the preceeding question has an affirmative answer. Let $R_{\mathrm{TOP}}(G)$ denote the quotient of the real representation $\mathrm{RO}(G)$ obtained by identifying all topologically equivalent representations. $R_{\mathrm{TOP}}(G)$ can be described directly as the free abelian group on the representation of G, modulo the subgroup generated by the elements $(\varrho_1 + \varrho_2 - \eta)$, where $\eta \sim_t \varrho_1 + \varrho_2$.

For s odd, let $K_s = Q\big(\exp(2\pi i/s)\big)$, R be the real part of the extension field of the rationals Q by the sth roots. Let $R_{K_s}(G)$ be the representation ring over K_s and let α be the composite

$$R_{K_s}(G) \to \mathrm{RO}(G) \to R_{\mathrm{TOP}}(G).$$

THEOREM 2.2. *Let G be a group of order $2^r s$, s odd. Then α induces an isomorphism*

$$R_{K_s}(G) \otimes Z[1/2] \cong R_{\mathrm{TOP}}(G) \otimes Z[1/2].$$

For G a 2-group, this result can be called a *topological rationality principle* for real representations, viz. $R_{\mathrm{TOP}}(G) \otimes Z[1/2] \cong R_Q(G) \otimes Z[1/2]$. For G of odd order, this theorem contains the results of Madsen–Rothenberg and Hsiang–Pardon. For the case of $|G|$ divisible by at least two odd primes, their results are used in the proof, and Schultz' theorem is used for $|G| = 2^r p^t$, p a prime, $p \neq 2$.

To conclude this section, here are two corollaries on which groups possess topologically equivalent representations that are not linearly equivalent.

COROLLARY 2.2.1. *Let G be a 2-group. Then the following are equivalent:*

(i) *Topologically similar representations of G are linearly similar;*

(ii) *For all $g \in G$, g is conjugate to g^a for some $a \equiv \pm 3$* (mod 8).

COROLLARY 2.2.2. *Let G be a finite group. Then the following are equivalent:*

(i) *Same as* (i) *in Corollary* 2.2.1.

(ii) *The real part of each character of a complex representation of G takes values in K_s, $|G| = 2^r s$, s odd.*

It should be noted that $R_{\mathrm{TOP}}(G)$ definitely has torsion, even for $G = Z/8Z$. Cappell and I believe that the determinitation of the torsion may depend heavily on certain number-theoretic phenomena. At any rate, the above results tell us precisely which groups have inequivalent representations that are topologically similar.

§ 3. Smith equivalence

The Smith equivalence problem also has an early history of positive results. Atiyah and Bott proved it for actions of Z/pZ, p a prime, and quoted an argument of Milnor to extend the result to actions that are semi-free (i.e. free away from the fixed point set, which is always the case for Z/pZ). Sanchez gives an argument that actually proves the conjecture, as stated in this paper, for all groups of *odd* order. One justification for the hypothesis of "Smith type" in the conjecture is that it's just what's needed for Sanchez' arguments to apply. Bredon gave some strong necessary conditions on Smith equivalent representation. For example, they show that if $|G| = 2^r$, then there is a number $\varphi(r)$ such that Smith equivalent representations of dimension more than $\varphi(r)$ are linearly similar. Bredon's theorem actually applies even when the action on the homotopy sphere is *not* of Smith type.

A representation of a group is pseudo-free if and only if the induced group action on the unit sphere of the representation space is free outside of a finite subset. Some of the results of Cappell and myself on Smith equivalence can be summed up in the following result:

THEOREM 3.1. *Let ϱ_1 and ϱ_2 be pseudo-free representations of a group G. Suppose that at least one of the following holds:* (i) *G is of order $4n$, n odd, or* (ii) $\dim \varrho_1 \leqslant 9$. *Then ϱ_1 and ϱ_2 are Smith equivalent* [6] *if and only if ϱ_1 and ϱ_2 are topologically similar.*

Cappell and I actually obtained a complete classification of Smith equivalence classes of pseudo-free representations of $G = Z_{2^r s}$, s odd, of

[6] It follows from Smith theory that any action on Σ with local representations ϱ_1 and α_2 will be free outside an $S^1 \subset \Sigma$.

dimension $\leqslant 9$ or $\geqslant \varphi(r)$. At least for pseudo-free representations, one can take $\varphi(r) \leqslant 2^r(2r)$. In this case as well, topological similarity and Smith equivalence will also be equivalent. Bredon's result shows that in general topological similarity does not imply Smith equivalence. However, the following is clearly indicated.

REVISED SMITH CONJECTURE. *Smith equivalent representations are topologically similar.*

This result can be proven using engulfing if, for $H \subset G$, $H \neq \{e\}$ Σ^H is actually a sphere of codimension at least three. Furthermore, Cappell and I proved:

THEOREM 3.2. *Let ϱ_1 and ϱ_2 be Smith equivalent representations of the finite group G. Then ϱ_1 and ϱ_2 are equal in*

$$R_{\mathrm{TOP}}(G) \otimes Z[1/2].$$

Without the hypothesis in the definition of "Smith equivalent" that Σ^H be discrete or connected for $H \subset G$, the Revised Conjecture, as well as Theorem 3.2, will definitely be false. This is illustrated by some examples of Petrie. If G is a group of *odd* order, with at least three non-cyclic Sylow subgroups, then Petrie's examples give actions of G on homotopy spheres Σ with $\Sigma^G = \{x, y\}$, and with the local representations of G at x and y not even topologically similar.[7] In his examples, Σ^H often has several components not all zero-dimensional, for some H.

The approach of Cappell and myself to the topological similarity problem applies to the Smith problem in the following way: an h-cobordism between unoriented interval bundles over the appropriate lens spaces is to be used as the quotient space of the action of the group on the complement in Σ of the singular set (i.e. the non-free set). Unlike the topological similarity problem, however, the Smith problem requires that the h-cobordism be both smooth *and* relative the boundary. This leads to some further homotopy theoretic conditions on Smith equivalent pseudo-free representations. More precisely, if ϱ_1 and ϱ_2 are pseudo-free representations of $Z_{2^a b}$, b odd, then there is an invariant $\varphi(\varrho_1, \varrho_2)$ in a quotient $[L \cup c\tilde{L}, G/O_{(2)}]$, whose vanishing, along with topological similarity, is necessary and sufficient for ϱ_1 and ϱ_2 to be Smith equivalent. (See [16], Theorem 1.) Here L is a suitable lens space, $\pi_1 L = Z_{2^a}$, and $\tilde{L}$ is its double cover. We show that $\varphi(\varrho_1, \varrho_2) = 0$ above dimension $\varphi(a)$ and in dimension $\leqslant 9$. (In fact, conditions like Bredon's can be derived from $\varphi(\varrho_1, \varrho_2) = 0$.) If $a \leqslant 2$, it is easy to see directly that $\varphi(\varrho_1, \varrho_2) = \varphi(\varrho_1 | Z_{2^a}, \varrho_2 | Z_{2^a}) = 0$.

[7] Actually not equal in $R_{\mathrm{TOP}}(G) \times Z[1/2]$ either.

Quite recently (1982), workers in Petrie's G-surgery theory have made further contributions on Smith equivalence. Dovermann has obtained some even dimensional Smith equivalent representations. In his 1982 thesis, Alan Siegel gave some examples of Smith equivalent pseudo-free representations of Z_{2^r}, in dimensions above 9. Petrie himself has also announced, (at least in lectures), results going beyond the pseudo-free case.

Finally, let me return, as promised, to the work of Hsiang–Pardon on topological similarity. Given a topological similarity $\varphi: (R^n, 0)\to(R^n, 0)$ of representations ϱ_1 and ϱ_2, one can use φ to glue together two copies of R^n along $R^n-\{0\}$. This yields a "topological Smith equivalence" of ϱ_1 with ϱ_2. Roughly speaking, the result of Hsiang and Pardon can be formulated as follows: If ϱ_1 and ϱ_2 are topologically similar representations of a cyclic group G, without fixed points on the unit sphere, and if all the restrictions to subgroups and restrictions to Σ^H for $H \neq \{e\}$ are linearly similar, then ϱ_1 and ϱ_2 are the local representations at isolated fixed points of a tame *piecewise linear action* of G on a sphere Σ. To obtain this result, they apply, in a careful and detailed way, the theory of Anderson–Hsiang on P. L. structures on stratified spaces.

For groups of odd order (and some others as well), Hsiang and Pardon then derive enough of an Atiyah–Singer theorem in the tame P. L. category to argue as Atiyah–Bott and Sanchez did in the smooth category for the Smith problem. (Their construction actually seems to give a G-sphere Σ of Smith type. However, they use the hypothesis of linear similarity on subgroups and fixed subspaces Σ^H directly in conjunction with arguments of the type of Sanchez.)

§ 4. Relation to number theory

We return to the condition of 2.2,

$$\Delta_A/\Delta_B = \gamma\big(T^\varepsilon\mu(T)\mu(-T)^{-1}\big), \quad \varepsilon = 0 \text{ or } 2q.$$

Obviously, this equation implies that the image in $Z[Z_{2q}]$ of Δ_A/Δ_B under the transfer $\big(f(T)\to f(T)f(-T)\big)$ is trivial. Let $\beta_1, \ldots, \beta_m$ be the eigenvalues of B as preceding 2.2 for A, and let $\beta_j s_j \equiv 1 \pmod{4q}$, and let α_i, r_i be as in the definition of Δ_A (just preceding 2.2). Then the usual argument with the Franz independence lemma (see [26]) implies that, after re-indexing, we have $r_i = s_i$ (modulo $2q$). The condition of 2.2 then becomes, after a suitable further re-indexing,

$$\prod_{i=1}^{k}(T^{s_i+2q}-1)/(T^{s_i}-1) = \gamma\big(T^{s}u(T)u(-T)^{-1}\big), \quad 1 \leqslant k \leqslant m. \tag{4.1}$$

If a is relatively prime to $4q$, let

$$f_a\colon Z/4qZ \to \{0, 1\}$$

be the function with $f_a(n)$, $0 \leqslant n < 4q$, having the value 1 if the least non-negative residue mod $4q$ of an is greater than zero and less than $2q$, and 0 otherwise. Then Cappell and I showed:

Theorem (4.1) *implies that k is even and*

$$(k/2)(f_1 + f_{2q+(-1)^q}) + \sum_{j=1}^{k} f_{s_j} \equiv 0.$$

Further, this condition is equivalent to (4.1) *if q is odd or $4q$ is a tempered number.*

The functions f_a satisfy the relations

$$f_a + f_{2q+(-1)^q a} = f_1 + f_{2q+(-1)^{2q}}. \tag{4.2}$$

A number is called *tempered* if all relations satisfied by the f_a are consequences of these. Examples of tempered numbers are 2^r, any r, and $4q$ for $q < 29$, q odd.

Now let q be *odd*, and let $G = G(q)$ be the Galois group of the field $Q(\zeta)$ over Q, ζ a prime qth root of unity. Let $\mathscr{S} \subset Z[G]$ be the Sticklberger ideal ([33], [22], [25]), and let $\mathscr{S}^-$ be the (-1)-eigenspace of the map induced by complex conjugation. Let $\delta\colon Z[G] \to Z_2[G]$ be reduction mod 2. For $(a, q) = 1$, let $\sigma_a \in G(q)$ be the element with $\sigma_a(\zeta) = \zeta^a$. Then in [18], Cappell and I prove essentially the following:

THEOREM. *The quotient of the relations among the functions $\{f_a \mid (a, 4q) = 1\}$ by the relations of the form* (4.2) *is isomorphic* (*as a $Z_2[G]$-module*) *to* $\mathrm{Ann}(\delta\mathscr{S}^-)/(\sigma_1 + \sigma_{-1})$.

A further lemma in [18] shows that $\mathrm{Ann}(\delta\mathscr{S}^-)/(\sigma_1 + \sigma_{-1})$ vanishes if and only if the inclusion $\mathscr{S}^- \subset (\sigma_1 + \sigma_{-1})$ is an equality. A theorem of Iwasawa for q a prime power and Sinnott in general says that if q is divisible by at most two primes, then the index of $\mathscr{S}^-$ in $Z[G]^-$ is precisely $h^-(q)$, the "first factor" of the class number $h(q)$ of q. The number $h^-(q)$ is the order of the (-1)-eigenspace of the involution on the ideal class group of $Q(\zeta)$ over Q induced by complex conjugation. Thus we have

THEOREM. *Let q have at most two prime divisors. Then $4q$ is tempered if and only if $h^-(q)$ is odd.*

If q has at least three prime divisors, Sinnott shows that the index of $\mathscr{S}^-$ in $Z[G]^-$ is even. So we have

THEOREM. *If q has at least three prime divisors, then $4q$ is not tempered.*

Here is one of our main geometric applications of these results.

THEOREM. *For q odd, the following are equivalent.*

(1) *Eigenvalues of topologically equivalent pseudo-free representations of Z_{4q} appear with the same multiplicities* mod 2.

(2) *Same as* (1) *but with "Smith equivalent" in place of "topologically equivalent".*

(3) *q has at most two prime factors and $h^-(q)$ is odd.*

Thus, if q is odd and (3) is valid, all topologically similar or Smith equivalent pairs, in the pseudo-free case, are given as in 2.2.1 above. When (3) does not hold, there remains the interesting question to describe the possible topologically equivalent representations in terms of the number theory of cyclotomic extensions. For example, what is the least dim $\lambda(q)$ in which pseudo-free representations can be topologically similar (or Smith equivalent), with eigenvalues with different multiplicities mod 2.

References

[1] Anderson D. W., The real K-theory of classifying spaces, *Proc. National Acad. Sciences USA* **51** (1964), pp. 634–636.

[2] Anderson D. R. and Hsiang W.-C., The functors K_{-i} and pseudoisotopies of polyhedra, *Ann. of Math.* **105** (1977), pp. 201–233.

[3] Anderson D. R. and Hsiang W.-C., Extending combinatorial P. L. structures on stratified spaces, *Invent. Math.* **32** (1976), pp. 176–204; II, *Trans. AMS* **260** (1980), pp. 223–253.

[4] Atiyah M. F., Characters and the cohomology of finite groups, Institute des Hautes Études Scientifiques, *Publications Mathématiques* **9** (1961), pp. 24–64.

[5] Atiyah M. F. and Bott R., A Lefshetz fixed point formula for elliptic complexes: II Applications, *Ann. of Math.* **88** (1968), pp. 451–491.

[6] Atiyah M. F. and Segal G. B., Equivariant K-theory and completion, *J. Diff. Geom.* **3** (1969), pp. 1–18.

[7] Atiyah M. F. and Singer I., The index of elliptic operators: *II Ann. of Math.* **87** (1968), pp. 546–604.

[8] Bredon G., Representations at fixed points of smooth actions of compact groups, *Ann. of Math.* **89** (1969), pp. 512–532.
[9] Bredon G., *Introduction to compact transformation groups*, Academic Press, New York, 1972.
[10] Browder W., *Surgery of simply-connected manifolds*, Springer-Verlag, New York, 1972.
[11] Cappell S. E. and Shaneson J. L., Linear Algebra and Topology, *Bull. A.M.S. (New Series)* **1** (1979), pp. 685–687.
[12] Cappell S. E. and Shaneson J. L., Non-linear similarity of matrices, *Bull. A.M.S. (New Series)* **1** (1979), pp. 899–902.
[13] Cappell S. E. and Shaneson J. L., Fixed points of periodic smooth maps, *Proc. Nat'l. Acad. of Sci. USA* **77** (1980), pp. 5052–5054.
[14] Cappell S. E. and Shaneson J. L., Non-linear similarity, *Annals of Math.* **113** (1981), pp. 315–355.
[15] Cappell S. E. and Shaneson J. L., The topology of linear representations of groups and subgroups, *Amer. J. of Math.* **104** (1982), pp. 773–778.
[16] Cappell S. E. and Shaneson J. L., Fixed points of periodic differentiable maps, *Inv. Math.* **68** (1982), pp. 1–19.
[17] Cappell S. E. and Shaneson J. L., The topological rationality of linear representations, Institut des Hautes Études Scientifiques, *Publications Mathématiques* **56** (1983), pp. 309–336.
[18] Cappell S. E. and Shaneson J. L., Class numbers and periodic smooth maps, *Commentarii Math. Helv.*, to appear.
[19] Dovermann H., *Even dimensional s-Smith equivalent representations*, preprint, 1982.
[20] Hsiang W.-C. and Pardon W., When are topologically equivalent orthogonal transformations linearly equivalent, *Inv. Math.* **68** (1982), pp. 275–316.
[21] Illman S., Representations at fixed points of actions of finite groups on spheres, *Canadian Math. Soc. Conference Proceedings* **2** (1982), pp. 135–155.
[22] Iwasawa K., A class number formula for cyclotomic fields, *Ann. of Math.* **76** (1962), pp. 171–179.
[23] Kirby R. and Siebenmann L. C., *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Ann. of Math. Studies, Princeton Univ. Press, 1977.
[24] Kuiper N. and Robbin J. W., Topological classification of linear endomorphisms, *Inv. Math.* **19** (1973), pp. 83–106.
[25] Lang S., *Cyclotomic fields*, Springer-Verlag, New York, 1978.
[26] Milnor J., Whitehead torsion, *Bull. A.M.S.* **72** (1966), pp. 358–426.
[27] Morgan J. and Sullivan D., The transversality characteristic class and linking cycles in surgery theory, *Ann. of Math.* **99** (1974), pp. 463–544.
[28] Madsen I. and Rothenberg M., Classifying G-spheres, *Bull. A.M.S. (New Series)* **7** (1982), pp. 223–226.
[29] Petrie T., *Three theorems in transformation groups*, Lecture Notes in Math. **763**, Springer-Verlag (1979).
[30] Rothenberg M., Torsion invariants and finite transformation groups, *Proc. Symp. in Pure Math.* **32** (1978), pp. 267–311.
[31] Sanchez C., Actions of groups of odd order on compact, orientable manifolds, *Proc. Amer. Math. Soc.* **54** (1976), pp. 445–448.

[32] Schultz R., On the topological classification of linear representations, *Topology* **16** (1977), pp. 263–270.
[33] Sinnott W., On the Stickelberger ideal and the circular units of a cyclotomic field, *Ann. of Math.* **108** (1978), pp. 107–134.
[34] Smith P. A., New results and dd problems in transformation groups, *Bull. A. M. S.* **66** (1960), pp. 401–415.
[35] Wall C. T. C., *Surgery on compact manifolds*, Academic Press, New York 1970.

**Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa**

A. BEILINSON

Localization of Representations of Reductive Lie Algebras

In this report I shall sketch the algebro-geometrical viewpoint on the representations of real reductive groups. One will see that representation theory deals with things quite familiar to algebraic geometers: namely with the geometry of Schubert-like varietes. As for applications, I shall confine myself to the following ones: a classification of irreducible representations [1], [13], the Kazhdan–Lusztig character formulas [7], [10] and the structure of Jantzen's filtration (the degeneration of series of representations). The last two subjects are based on arithmetical considerations (the theory of mixed perversed sheaves [3], [4]).

All the schemes considered will be over a fixed algebraically closed ground field k of char 0.

A. Affine spaces and localization

Let X be a scheme.

DEFINITION. An *$\mathcal{O}_X$-ring* is a sheaf $\mathcal{R}$ of rings on X, together with a ring morphism $\mathcal{O}_X \to \mathcal{R}$ such that $\mathcal{R}$ is quasicoherent as a left $\mathcal{O}_X$-module. For an $\mathcal{O}_X$-ring $\mathcal{R}$, an $\mathcal{R}$-module is a sheaf of left $\mathcal{R}$-modules, quasicoherent as a sheaf of $\mathcal{O}_X$-modules. ∎

Denote by $\mathcal{R}$-mod the category of $\mathcal{R}$-modules. Put $R := \Gamma(X, \mathcal{R})$. There are natural adjoint functors $\mathcal{R}\text{-mod} \underset{\Delta}{\overset{\Gamma}{\rightleftarrows}} R\text{-mod}$: $\Gamma(\mathcal{M}) := \Gamma(X, \mathcal{M})$, $\Delta(N) = \mathcal{R} \underset{R}{\otimes} N$, and also corresponding derived functors $R\Gamma$ and $L\Delta$.

DEFINITION. We shall say that X is *$\mathcal{R}$-affine* if Γ and Δ are (mutually inverse) equivalences of categories; and that X is *$\mathcal{R}$-affine in the sense of derived categories of amplitude $\leqslant n$* if $R\Gamma$ is equivalence of derived categories of amplitude $\leqslant n$. ∎

If for a ring R there exist some $(X, \mathscr{R})$ such that $R = \Gamma(X, \mathscr{R})$ and X is $\mathscr{R}$-affine, then one may study R-modules by local methods as sheaves over X; in this situation we call the sheaf $\Delta(N)$ the *localization* of an R-module N.

Here is a criterion for $\mathscr{R}$-affinity. Any $\mathscr{R}$-module $\mathscr{M}$ is generated by global sections and $H^i(X, \mathscr{M}) = 0$ for $i > 0$.

The $\mathcal{O}_X$-rings that we shall consider are the rings of differential operators $\mathscr{D}_X$ on some smooth X, or slightly more general rings introduced in the next section. Here is a somewhat striking example (see Section C): any flag space X (e.g. $X = \boldsymbol{P}^N$) is $\mathscr{D}_X$-affine.

B. Twisted rings of differential operators

Let X be a smooth variety.

DEFINITION. An $\mathcal{O}_X$-ring is called a *ring of twisted differential operators* (tdo for short) if it is locally isomorphic (on X), as an $\mathcal{O}_X$-ring, to $\mathscr{D}_X$. If A is any commutative k-algebra, then an A-tdo is a sheaf of A-algebras, and also an $\mathcal{O}_X$-ring, locally isomorphic to $\mathscr{D}_X \underset{k}{\otimes} A$. ■

Remark. A-tdo is just a family of tdo's, parametrized by Spec A.

As the automorphisms of $\mathcal{O}_X$-ring $\mathscr{D}_X$ are exactly closed 1-forms on X (a 1-form ω corresponds to an automorphism $\dfrac{\partial}{\partial x} \to \dfrac{\partial}{\partial x} + \omega\left(\dfrac{\partial}{\partial x}\right)$), one may identify tdo's with $\Omega_X^{1,\mathrm{cl}}$-torsors (and A-tdo's with $A \otimes \Omega_X^{1\mathrm{cl}}$-torsors). In particular tdo's form a "linear k-space". If χ is a torsor, we denote the corresponding tdo by $\mathscr{D}_X^\chi$. The set of isomorphic classes of tdo is $H^1(X, \Omega_X^{1,\mathrm{cl}})$, for a proper X it is the k-subspace of $H^2_{DR}(X)$ generated by algebraic cycles.

Examples. (a) If $\mathscr{L}$ is any invertible $\mathcal{O}_X$-sheaf, then the sheaf $\mathscr{D}_{X,\mathscr{L}}$ of differential operators on $\mathscr{L}$ is a tdo. One has $\mathscr{D}_{X,\mathscr{L}} = \mathscr{D}_X^{d\log \mathscr{L}}$.

(b) If $\mathscr{D}$ is tdo, then $\mathscr{D}^0$ — a ring with inverted multiplication — is also a tdo. If χ is an $\Omega_X^{1,\mathrm{cl}}$-torsor, put $\chi^0 := d\log \Omega - \chi$, where Ω is the sheaf of volume forms $\det \Omega_X^1$. One has $(\mathscr{D}_X^\chi)^0 = \mathscr{D}_X^{\chi^0}$, as Ω has the canonical structure of a *right* $\mathscr{D}_X$-module.

C. Main construction

Let G be a connected reductive group over k, $\mathscr{G} := \mathrm{Lie}\, G$ its Lie algebra, $U = U(\mathscr{G})$ the universal enveloping algebra, and Z the centre of U. Let X be the flag manifold of G (the space of its Borel subgroups), and T_X the sheaf of vector fields on X. For $x \in X$ let $B_x \supset N_x$ be the corresponding Borel subgroup and its radical, $H := B_x/N_x$ the Cartan group (it does not depend on x), and $b_x \supset n_x$ and h the corresponding Lie algebras. There is a Harish-Chandra map $Z \subset S(h)$ which identifies Z with the space of W-invariant polynomials on h^* (where the Weyl group W acts fixing ϱ — the half-sum of positive roots), let Θ: Spec $S(h) \to$ Spec Z be the corresponding map. Let $\tilde{U} := U \otimes_Z S(h)$ be the extended universal enveloping algebra. It is clear that the centre of $\tilde{U}$ is $S(h)$.

Now we pass to the construction. The group G acts on X and one has the corresponding Lie algebra map α: $\mathscr{G} \to T_X$. Define an $\mathcal{O}_X$-ring structure on $U_X := \mathcal{O}_X \otimes_k U$ by demanding that $[A, f] = \alpha(A)f$ for $A \in \mathscr{G} \subset U, f \in \mathcal{O}_X$ and that the multiplication on $U \subset U_X$ should coincide with the usual one. Consider the induced Lie algebra structure on $\mathscr{G}_X := \mathcal{O}_X \otimes \mathscr{G} \subset U_X$, and put $\mathscr{B}_X = \ker(\alpha\colon \mathscr{G}_X \to T_X) = \{\xi \in \mathscr{G}_X\colon \forall x \in X\, \xi(x) \in b_x\}$, $\mathscr{N}_X = [\mathscr{B}_X, \mathscr{B}_X] = \{\xi \in \mathscr{G}_X\colon \forall x \in X \ \ \xi(x) \in n_x\}$. These are ideals in $\mathscr{G}_X$. Put $\tilde{\mathscr{D}} := U_X / U_X \cdot \mathscr{N}_X$. One has obvious maps $U \to \Gamma(X, \tilde{\mathscr{D}})$, and $S(h) \to \Gamma(X, \tilde{\mathscr{D}})$ (as $\mathcal{O}_X \otimes h = \mathscr{B}_X/\mathscr{N}_X \subset U_X$). It is easy to see that $S(h)$ is mapped onto the centre of $\tilde{\mathscr{D}}$, and both these maps coincide on Z. So they define the map $\tilde{U} \to \Gamma(X, \tilde{\mathscr{D}})$.

LEMMA. (a) *This map is an isomorphism*: $\tilde{U} = \Gamma(X, \tilde{\mathscr{D}})$.
(b) $\tilde{\mathscr{D}}$ *is an* $S(h)$*-tdo on* X. ∎

To apply this to the study of $\mathscr{G}$-modules we need to verify the affinity of the picture. For simplicity we confine ourselves to Z-finite $\mathscr{G}$-modules, i.e., $\mathscr{G}$-modules annihilated by some ideal of finite codimension in Z. Let $\chi \in h^*$ be a character, and $m_\chi \subset S(h)$ the corresponding maximal ideal. Put $\widehat{S(h)}_\chi := \varprojlim S(h)/m_\chi^n$, $\mathscr{D}_\chi := \tilde{D}/m_\chi \tilde{\mathscr{D}}$, $\hat{\mathscr{D}}_\chi := \tilde{\mathscr{D}} \otimes_{S(h)} \widehat{S(h)}_\chi$. It is clear that $\mathscr{D}_\chi$ is a tdo (and $\hat{\mathscr{D}}_\chi$ is an $\widehat{S(h)}_\chi$-tdo), and one has $\Gamma(X, \mathscr{D}_\chi) =: U_\chi = \tilde{U}/m_\chi \tilde{U}, \Gamma(X, \hat{\mathscr{D}}_\chi) =: \hat{U}_\chi = \hat{U} \otimes \widehat{S(f)}_\chi$.

Remark. If χ is integral, i.e., originates in the character of H, then $\mathscr{D}_\chi$ is a sheaf of differential operators on a corresponding invertible sheaf on X. In particular $\mathscr{D}_0 = \mathscr{D}_X$.

Now suppose χ to be regular. Then Θ: Spec $S(h) \to$ Spec Z is étale at χ and so $\widehat{S(h)}_\chi = \hat{Z}_{\Theta(\chi)}, \hat{U}_\chi = \hat{U}_{\Theta(\chi)} := U \otimes_Z \hat{Z}_{\Theta(\chi)}, U_\chi = U/m_{\Theta(\chi)} U$. This means that $\hat{U}_\chi(\hat{U}_{\Theta(\chi)})$-modules are just $\mathscr{G}$-modules with a (generalized) central character $\Theta(\chi)$. Denote by Γ_χ, Δ_χ the corresponding functors $\overset{(\wedge)}{D}_\chi\text{-mod} \rightleftarrows \overset{(\wedge)}{U}_{\Theta(\chi)}-\text{mod}$.

Recall that $\chi \in h^*$ is dominant if for any simple positive coroot l one has $\chi(l) \neq 0, -1, \ldots$

THEOREM [1], [2]. *Let χ be a regular weight. If χ is dominant then X is $\mathscr{D}_\chi$- and $\hat{\mathscr{D}}_\chi$-affine. If $\chi = w\chi_0$ where χ_0 is dominant, then X is $\mathscr{D}_\chi$- and $\hat{\mathscr{D}}_\chi$-affine in the sense of the derived categories, of amplitude $\leqslant$ the length of w.* ■

So $\mathscr{G}$-modules with a regular central character θ are just $\mathscr{D}_\chi$-modules for dominant χ such that $\Theta(\chi) = \theta$.

Remark. Let M be a U_θ-module and $x \in X$. One knows that the spectrum of the natural action of h on $H.(n_x, M)$ is contained in $\Theta^{-1}(\theta)$: put $H. = \bigoplus_{\chi\in\Theta^{-1}(\theta)} H._\chi$. The definition implies that $H.(n_x, M)_\chi = \mathrm{Tor.}^{\mathcal{O}_x}(k_x, L\Delta_\chi M)$ — the fibre at x of $L\Delta_\chi M$ in the sense of $\mathcal{O}_X$-modules. So the theorem is a generalization (to arbitrary $\mathscr{G}$-modules) of the highest-weight and Borel–Weil–Bott theory of finite-dimensional representations.

D. Functorial properties of $\mathscr{D}$-modules

Here we recall some basic functors on $\mathscr{D}$-modules (following Bernstein, Kashiwara, ...) and show what these functors mean for representations.

D1. Translation. Let $\mathscr{L}$ be an invertible $\mathcal{O}_X$-sheaf and χ any torsor; then $\mathscr{D}_X^\chi$-mod is canonically equivalent to $\mathscr{D}_{X,\mathscr{L}}^\chi$ $(= \mathscr{D}_X^{\chi+d\log\mathscr{L}})$-mod: one transforms the $\mathscr{D}^\chi$-module $\mathscr{M}$ to $\mathscr{L}\underset{\mathcal{O}_X}{\otimes}\mathscr{M}$ with a canonical $D_{X,\mathscr{L}}^\chi$-action. So for any regular dominant $\chi_1, \chi_2 \in h^*$ such that $\chi_1 - \chi_2$ is integral one has a canonical equivalence between the categories $\overset{(\wedge)}{U}_{\theta(\chi_i)}$-mod; this is the Bernstein–Gelfand translation principle.

D2. Action of correspondences. Let $f\colon Y \to X$ be a morphism of smooth varietes, and $\mathscr{D}_{(X)} = \mathscr{D}_X^\chi$ be a tdo on X; then one has a tdo $\mathscr{D}_{(Y)} := D^{f^{-1}(\chi)}$ on Y. Define exact functors $f^!\colon \mathscr{D}(\mathscr{D}_{(X)}\text{-mod}) \to \mathscr{D}(\mathscr{D}_{(Y)}\text{-mod})$ and $f_*\colon \mathscr{D}(\mathscr{D}_{(Y)}\text{-mod}) \to \mathscr{D}(\mathscr{D}_{(X)}\text{-mod})$ between derived categories as follows. First, for any $\mathscr{D}_{(X)}$-module $\mathscr{M}$ there is a natural $\mathscr{D}_{(Y)}$ action on $f^+(\mathscr{M}) := \mathcal{O}_Y \underset{\mathcal{O}_X}{\otimes} \mathscr{M}$ — the inverse image of $\mathscr{M}$ in the sense of $\mathcal{O}$-modules. Let $Lf^+\colon \mathscr{D}(\mathscr{D}_{(X)}\text{-mod})$

$\to \mathscr{D}(\mathscr{D}_{(Y)}\text{-mod})$ be a derived functor of f^+; put $f^! := Lf^+[\dim Y - \dim X]$. To define f_* one uses $(\mathscr{D}_{(Y)} - f^{\cdot}(\mathscr{D}_{(X)}))$-bimodule structure on $f^+(\mathscr{D}_{(X)})$: put $f_*(\mathscr{N}) = Rf_{\cdot}(\mathscr{N} \overset{L}{\underset{\mathscr{D}(Y)}{\otimes}} f^+(\mathscr{D}_{(X)}))$.

Examples. (a) If f is a closed imbedding, then f_* and $f^!$ induce equivalence between $\mathscr{D}_{(Y)}$-mod and the category of $\mathscr{D}_{(X)}$-modules supported on Y.

(b) If f is smooth, then $f_*(\mathscr{M})$ is the relative de Rham complex with coefficients in $\mathscr{M}$, shifted by the relative dimension of f.

Remark. These functors may be unified by considering correspondences. Namely, consider the diagram

$$\begin{array}{ccc} & Z & \\ {\scriptstyle \pi_Y}\swarrow & & \searrow{\scriptstyle \pi_X} \\ Y & & X \end{array}$$

where Z is also smooth, and two torsors, χ_X on X and χ_Y on Y, with a fixed isomorphism $\pi_Y^{-1}(\chi_Y^0) = \pi_X^{-1}(\chi_X^0)$. Let $Z_*: \mathscr{D}(\mathscr{D}_Y^{\chi_Y}\text{-mod}) \to \mathscr{D}(\mathscr{D}_X^{\chi_X}\text{-mod})$ be $\pi_{X*}\pi_Y^!$.

Now let us return to representations. Let X be as in Section C and $w \in W$. We have the Bruhat–Hecke correspondence $N_w = \{(x, x') \in X \times X$ such that $(b_x, b_{x'})$ are in relative position $w\}$.

THEOREM [2]. *Let χ be dominant regular. Then* $(N_w)_* L\Delta_\chi = L\Delta_{w(\chi)}$. ∎

So the action of the Bruhat–Hecke correspondences on $\mathscr{D}$-modules (the intertwining functors of [2]) corresponds to the action of W on $\tilde{U}$. These correspondences play a very important role in the Kazhdan–Lusztig theory (of Section H).

D3. Duality. Let $\mathscr{D}$ be a tdo. One says that $\mathscr{D}$-module is *coherent* if it is locally finitely generated; let $\mathscr{D}$-modc be the category of coherent modules. Consider $\mathscr{D}$ as $(\mathscr{D}, \mathscr{D})$-bimodule; it defines a natural functor $*$: $\mathscr{D}(\mathscr{D}\text{-modc})^0 \to \mathscr{D}(\mathscr{D}^0\text{-modc})$ by the formula $*\mathscr{M} := R\,\mathrm{Hom}(\mathscr{M}, \mathscr{D}[\dim X])$; one has $** = \mathrm{id}$. Coherent modules correspond to finitely generated representations, and $*$ corresponds to the functor $*M := R\mathrm{Hom}(M, U_\theta[\dim X])$.

Example. One says that a $\mathscr{D}$-module is smooth if it is coherent as a sheaf of $\mathcal{O}_X$-modules, or, equivalently, if after a (local) isomorphism $\mathscr{D} \simeq \mathscr{D}_X$ it becomes a sheaf of sections of a bundle with integrable connection. For smooth $\mathscr{M}$ we have $*\mathscr{M} = \mathrm{Hom}_{\mathcal{O}_X}(\mathscr{M}, \Omega_X)$ with an obvious $\mathscr{D}^0$-module structure.

E. Holonomic modules and the Harish-Chandra modules

The holonomic $\mathscr{D}$-modules are those that are smooth along a certain stratification. Let me introduce one important construction before giving the exact definition. Consider the affine locally closed imbedding $i: Y \hookrightarrow X$, where Y is smooth, and a smooth $\mathscr{D}_{(Y)}$-module $\mathscr{M}$. Then $i_*\mathscr{M}$ is a coherent $\mathscr{D}_{(X)}$-module. Put $i_! \mathscr{M} := *i_* * \mathscr{M}$; we have $i^! i_! \mathscr{M} = i^! i_* \mathscr{M} = \mathscr{M}$ and there is a unique morphism φ: $i_!\mathscr{M} \to i_*\mathscr{M}$ such that $i^!(\varphi) = \mathrm{id}_{\mathscr{M}}$. Denote $\mathrm{Im}\,\varphi$ by $i_{!*}\mathscr{M}$. If $\mathscr{M}$ is irreducible, then $i_{!*}\mathscr{M}$ is the unique irreducible submodule of $i_*\mathscr{M}$ (and the unique irreducible quotient of $i_!\mathscr{M}$). In this situation the modules $i_*\mathscr{M}$, $i_!\mathscr{M}$ are called *standard modules* and $i_{!*}\mathscr{M}$ is called the *irreducible module* corresponding to $(Y, \mathscr{M})$.

By definition a $\mathscr{D}$-module $\mathscr{M}$ is holonomic if it has a finite length and all of his Jordan–Hölder components are of the type constructed above. One says that holonomic $\mathscr{M}$ (on compact X in the twisted case) has regular singularities (RS's for short) if all its components originate in bundles with regular singularities at infinity. The basic property of holonomic modules is that the corresponding derived category of complexes with holonomic cohomology is stable under the functors of the type $f^!, f_*$; if $\mathscr{M}$ is holonomic, then $*\mathscr{M}$ is also holonomic. The same applies for holonomic RS's.

Now return to $\mathscr{G}$-modules. The representation theorists claim that the representations that happen in nature are $(\mathscr{G}, K)$-modules for a certain (algebraic) subgroup $K \subset G$ (cf. [13]).

Roughly speaking, a $(\mathscr{G}, K)$-module N is $\mathscr{G}$-module s.t. the action of Lie K may be integrated to an algebraic representation of K. Indeed, one simply fixes the algebraic action of K on N with obvious compatibility conditions. It is easy to see that $(\mathscr{G}, K)$-modules correspond to $(\mathscr{D}, K)$-modules on X, i.e., to $\mathscr{D}$-modules with such an action of K that Lie K acts via the imbedding Lie $K \hookrightarrow \mathscr{D}$.

To get an interesting theory one needs sufficiently large subgroups K. Say that K is admissible if Lie K is transverse to some Borel subalgebra or, equivalently, if K acts on X with finitely many orbits. Fix an admissible K. It is easy to see that any coherent $(\mathscr{D}, K)$-module is smooth along the orbits of K and is holonomic RS. The irreducible $(\mathscr{D}, K)$-modules are in a 1–1 correspondence by means of the $i_{!*}$ construction, with irreducible smooth $(\mathscr{D}_{(Y)}, K)$-modules on different K-orbits Y, and these are defined by representations of the stabilizers of points.

Arrange this to get a classification of irreducible $(\mathscr{D}, K)$ (and so of $(\mathscr{G}, K)$)-modules. For any K-orbit Y put $H_Y := K \cap B_y/_{K \cap N_y} \subset H$ where

$y \in Y$ (note that H_Y does not depend on $y \in Y$); $h_Y := \mathrm{Lie} H_Y$, and thus H_Y is a product of the torus H_Y^0 and the finite abelian group H_Y/H_Y^0.

THEOREM [1], [13]. *For $\chi \in h^*$, the irreducible $(\mathscr{D}_\chi, K)$-modules are in a 1–1 correspondence with the set of pairs (Y, τ_Y), where Y is a K-orbit on X and τ_Y is the irreducible (h, H_Y)-module on which h acts by χ. If χ is regular dominant, this is also the classification of irreducible $(\mathscr{G}, K)$-modules with central character $\Theta(\chi)$.* ■

Remarks. (a) We see that the standard and irreducible modules that correspond to an orbit Y form $\dim(h/_{h_Y})$-parameter families (integrality condition on $\chi/_{h_Y}$). All the standard modules are irreducible for generic values of the parameter. If they are irreducible for any value of the parameters (or for some integral one) then Y is a closed orbit — this is the case of (generalized) discrete series. Some information on the non-closed orbits case will be presented in Section I. (In fact, to work clearly with the standard modules one has to suppose that for any orbit its imbedding into X is affine; this is the case in any example of the next remark.)

(b) The subgroups K usually considered are the fixed points of involutions of G (this correspond to the Harish-Chandra modules or representations of real reductive groups) or either N_x or B_x for certain $x \in X$ (the representations with highest weights); the second case may be reduced to the first (the representations of complex reductive groups). Such subgroups K are admissible. The standard modules for $K = N$ are just the Verma modules and the dual ones; in the Harish-Chandra situation the standard modules (and thus the classification) coincide with those of Langlands. This was proved by Vogan in [12]: he compared the above construction with that of Zuckerman (the cohomological parabolic induction). It would be interesting to use some analysis to compare it directly with that of Langlands. I understand that J. Bernstein has been working on this subject.

(c) The standard modules are considered as the simplest from the representation-theoretic point of view; in particular in Harish-Chandra's situation their characters are known. As both the standard and the irreducible modules form bases in the Grothendieck group of $(\mathscr{G}, K)$-modules, it is of importance to find the ones in terms of the others. And this is what the Kazhdan–Lusztig algorithm does.

F. Perverse sheaves

Now suppose that our ground field is $\boldsymbol{C}$. Let us describe the topological interpretation of holonomic RS $\mathscr{D}$-modules. In the case of smooth modules what follows is just the old result of Deligne, which claims that the local systems are the same as bundles with integrable connections having RS at infinity.

Let X be a smooth variety and $\mathscr{M}$ a $\mathscr{D}_X$-module. Denote by $\Omega(\mathscr{M})$ the analytic de Rham complex of $\mathscr{M}$: $\Omega(\mathscr{M}) := (\mathscr{M} \to \Omega^1 \underset{\mathcal{O}_X}{\otimes} \mathscr{M} \to \ldots)$, $\mathscr{M}$ being placed in degree $-\dim X$. This defines the functor Ω from the derived category of $\mathscr{D}_X$-modules to that of sheaves on X_{an}. Kashiwara proved that $\Omega(\mathscr{M})$ has a constructible cohomology for holonomic $\mathscr{M}$. And, due to Kashiwara, Kawai and Mebkhout, we have the following comparison theorem.

THEOREM. *The functor Ω induces an equivalence between the derived category of the complexes of $\mathscr{D}_X$-modules, having a holonomic RS cohomology, and that of the complexes of C-sheaves on X, having a constructible cohomology. This equivalence transforms $*$ to the Verdier duality, and $f_*, f^!$ to the functors of the same notation from the constructible sheaf theory.* ∎

Example. Return to the situation from the beginning of Section E; suppose that $\mathscr{M}$ has RS. Then $\Omega(\mathscr{M})$ is the local system, corresponding to $\mathscr{M}$, placed in degree $-\dim Y$. We have $\Omega(i_!\mathscr{M}) = i_!\Omega(\mathscr{M})$, the same applies to i_*, and $\Omega(i_{!*}\mathscr{M})$ is the Deligne–Goresky–MacPherson complex of $\Omega(\mathscr{M})$ on $\overline{Y}$, prolonged by 0 on $X \setminus \overline{Y}$ ([1], [5]). ∎

But what about $\mathscr{D}_X$-modules themselves? If $\mathscr{M}$ is a holonomic RS, then according to Kashiwara we have the following conditions on $\Omega(\mathscr{M})$: $\dim \operatorname{supp} H^i(\Omega(\mathscr{M})) \leqslant -i$ for any i, and the same applies to $*\Omega(\mathscr{M})$. Such constructible complexes are called *perverse sheaves* [3], [4], and the theorem implies that Ω induces the equivalence between the category of $\mathscr{D}_X$-modules holonomic RS and that of perverse C-sheaves.

Let me show how to deal with the tdo case. To be brief, consider the flag space only. There is a canonical H-torsor π: $\tilde{X} = G/N \to X = G/B$ over X ("the base affine space"). The ring $\pi_*(\mathcal{O}_{\tilde{X}})$ is graded by the weight lattice, and we have the corresponding gradation on $\pi_*\mathscr{D}_{\tilde{X}}$. The zero component of $\pi_*\mathscr{D}_{\tilde{X}}$ is just $\tilde{\mathscr{D}}$, so $\tilde{\mathscr{D}}$-modules are the same as graded $\pi_*\mathscr{D}_{\tilde{X}}$-modules — and so are $\mathscr{D}_{\tilde{X}}$-modules. This leads to the following perverse description of $\mathscr{D}_\chi$- and $\hat{\mathscr{D}}_\chi$-modules. A perverse sheaf $\mathscr{F}$ on $\tilde{X}$ is said to be *monodromic* if it is smooth along the fibres of π. For such $\mathscr{F}$ one has the monodromy representation (along the fibres) of the co-weight lattice $(= \pi_1(H))$ in $\operatorname{Aut} \mathscr{F}$ [11]. Then the category of (holonomic RS) $\mathscr{D}_\chi$-modules is equivalent to that of monodromic perverse C-sheaves on $\tilde{X}$ of monodromy $\exp\chi$ and $\hat{\mathscr{D}}_\chi$-modules are monodromic sheaves such that every eigenvalue of monodromy is $\exp\chi$.

Remark. The advantage of replacing holonomic modules by perverse sheaves is the possibility to use any coefficient ring (e.g., one has the Q_l-Harish-Chandra modules). Since Q_l-perverse sheaves has an étale meaning, this opens the door to arithmetics (certainly, there should be an arithmetical crystalline theory of $\mathscr{D}$-modules, but at a moment one has to use constructible sheaves).

G. Motivic language: mixed perverse sheaves

The yoga of motives claims that there should exist a fine category of motives over X such that any "natural" perverse sheaf on X or $\mathscr{D}_X$-module is a realization of a certain motive. Since the motivic theory has not appeared yet, one is forced to use its l-adic realization — the theory of mixed perverse sheaves [3], [4], based on the Weil conjectures proved by Deligne [6].

Here are a few properties of mixed sheaves. There is an abelian category $\mathscr{M}_{\text{mixed}}(X)$ of mixed perverse sheaves on X, together with canonical functor from $\mathscr{M}_{\text{mixed}}(X)$ to the category $\mathscr{M}(X)$ of Q_l-perverse sheaves. Any mixed perverse sheaf M has a canonical (finite) decreasing weight filtration $W.(M)$; any morphism in $\mathscr{M}_{\text{mixed}}$ is strongly compatible with the weight filtration. The object $\mathrm{Gr}_{W.}(M)$ is semi-simple, at least in $\mathscr{M}(X)$. There is Verdier's duality functor $*$ on $\mathscr{M}_{\text{mixed}}$, compatible with the one on $\mathscr{M}(X)$; one has $*W_i(M) = W_{-i}(*M)$. There is also the corresponding derived category of mixed sheaves, together with all the standard functors, compatible with the one on usual sheaves.

Thus, in this way we get the category of mixed Harish-Chandra modules (this category will be non-empty for rational χ, but, if that is so, then any irreducible module has a mixed structure).

Problem. Construct this category by representation theoretical means (any natural representation should get a mixed structure: and so the weight filtration with the properties above).

I shall mention two applications of this mixed category. The first is the Kazhdan–Lusztig algorithm (the starting point of all the things above) and the second one is the structure of Jantzen's filtration on standard modules.

H. The Kazhdan–Lusztig algorithm

What follows is only a very rough exposition of the basic ideas; the algorithm itself may be found in [7], [10] (the Verna modules case) and in [12], [13] (the general Harish-Chandra case, due to Vogan).

Suppose we are given a variety X stratified by strata X_j and also a number of irreducible local systems ($=$ smooth perverse sheaves) $\{V_j\}$ on any X_j. Suppose that every $i_j\colon X_j \hookrightarrow X$ is affine, and that irreducible components of any $i_!(V_j)$ are isomorphic to some $i_{j'!*}(V_{j'})$. Let $\mathscr{U}$ be a subgroup of the Grothendieck group of perverse sheaves on X generated by $\{i_{j!*}(V_j)\}$. It has two natural bases: the irreducible one $\{i_{!*}(V)\}$ and the standard one $\{i_!(V)\}$. The problem is to compute $\{i_{!*}\}$ via $\{i_!\}$ (cf. Section E, Remark c).

To do this, suppose that everything has arisen from a mixed situation (and so we have $\{V\}_{\mathrm{mixed}}, \mathscr{U}_{\mathrm{mixed}}, \ldots$), and that the set $\{V\}_{\mathrm{mixed}}$ is $*$-closed. Then $*$ acts on $\mathscr{U}_{\mathrm{mixed}}$. We have one extra structure on $\mathscr{U}_{\mathrm{mixed}}$ — namely the weight filtration $W_.$. The space W_j may be defined in terms of standard bases, as the one generated by all $i_!(V)$ such that V is of weight $\leqslant j$. Then, if the weight of V is j, then $i_{!*}(V)$ is the unique element of $\mathscr{U}_{\mathrm{mixed}}$, such that $i_{!*}(V) \in W_j \cap *W_{-j}$ and $i_{!*}(V) - i_!(V) \in W_{j-1}$. So to compute $\{i_{!*}\}$ in terms of $\{i_!\}$ it suffices to know the matrix of $*$ in $i_!$-bases.

To find $*$ in the representation-theoretic situation, one uses the action of the Hecke algebra (already appearing in the theorem of Section D2 — but now we need its mixed variant). The Hecke algebra $\mathscr{H}$ is the $\mathscr{U}_{\mathrm{mixed}}$-group related to the stratification of $X \times X$ by G-orbits (and the constant sheaves). The multiplication on $\mathscr{H}$ is the multiplication of correspondences. This algebra — the mixed variant of the group algebra of the Weyl group — may be given explicitly by generators and relations, and the $*$ operator on $\mathscr{H}$ is given by compatibility with multiplication and by an explicit formula on generators [7], [8]. Let us return to Harish-Chandra's modules. The Hecke algebra acts on the corresponding $\mathscr{U}_{\mathrm{mixed}}$ group, and the $*$ operator on $\mathscr{U}_{\mathrm{mixed}}$ is more or less determined by compatibility with this action and by the claim that $i_!(V) = i_{!*}(V)$ for closed orbits [9], [12].

I. The Jantzen's filtration

This filtration shows how the standard representations become reducible under the specialization of parameters.

Return to the situation of Section E. Fix an orbit Y, the character $\chi \in h^*$, integral with respect to Y, and the (h, H_Y)-module τ of the theorem quoted there. We have the corresponding standard modules $i_!(\tau)$, $i_*(\tau)$ and the canonical morphism $i_!(\tau) \to i_*(\tau)$, whose image is the irreducible $i_{!*}(\tau)$. Now we may vary τ in the family having $(f/h_Y)^* + \tau \subset f^*$ as parameters (see Remark (a) of Section E) to obtain the corresponding $\hat{\mathscr{D}}_\chi$-modules $i_!(\hat{\tau})$ and $i_*(\hat{\tau})$ together with the morphism $i_!(\hat{\tau}) \to i_*(\hat{\tau})$.

From now on we suppose that Lie $K = \mathscr{G}^\sigma$ for certain involution σ of $\mathscr{G}$ (see Remark (b) of Section E; the case $K = N_x$ is even simpler). Let C be the intersection of $(f/h_Y)^*$ with the positive cone of rational characters. If Y is not closed, then C is non-zero. Choose φ in the open cone C^0 and consider the one-parameter subfamily τ_φ of $\hat{\tau}$ that depends on $\chi + t\varphi$. One knows that $i_!(\tau_\varphi) \to i_*(\tau_\varphi)$ is an injection whose cokernel $i_*/i_!(\tau_\varphi)$ is of finite length.

Define the filtration $I^{(\varphi)}_{\cdot}$ on $i_*(\tau)/i_{!*}(\tau) = i_*/i_!(\tau_\varphi)/ti*/i!(\tau)(\tau_\varphi)$ by the formula $J^{(\varphi)}_n := [\mathrm{Ker}(t^n \in \mathrm{End}\,(i_*(\dot{\tau}_\varphi)/_{i!x}(\tau_\varphi))]\mathrm{mod}\, t$; this defines the filtration $J^{(\varphi)}_{\cdot}$ on $i_*(\tau)$ such that $i^{(\varphi)}_0 = i_{!*}(\tau)$.

This is Jantzen's filtration.

Remark. One may show, that the standard $\mathscr{G}$-module $i_!(\tau)$ is the K-finite dual to a certain $i_*(\tau')$; so Jantzen's filtration may be defined in terms of a "contravariant form".

THEOREM. *The filtration $I^{(\varphi)}$ coincides, up to a shift, with the weight filtration on the mixed perverse sheaf $i_*(\tau)$.* ■

COROLLARY. *The filtration $I^{(\varphi)}_{\cdot}$ does not depend on the choice of $\varphi \in C^0$. The module $\mathrm{gr}_{J_{\cdot}(\varphi)}(i_*(\tau))$ is a direct sum of irreducible ones. The multiplicities in terms of Jantzen's filtration are given by the Kazhdan–Lusztig–Vogan algorithm.* ■

To prove the theorem one has to identify the module $i_*/i_{!*}(\tau_\varphi)$ with certain sheaf of vanishing cycles, and then to use Gabber's purity theorem for vanishing cycles.

The theorem above, due to J. Bernstein and the present author, was conjectured (in the Verma modules case) by Brylinski; the corresponding numerical statement is the generalized Kazhdan–Lusztig conjecture of Gabber–Joseph and S. Gelfand–MacPherson.

References

[1] Beilinson A. and Bernstein J., Localisation de $\mathscr{G}$-modules, *C. R. Acad. Sc. Paris* **292**, pp. 15–18.

[2] Beilinson A. and Bernstein J., A Generalisation of a Theorem of Casselman, *Report on Utah Conference on Representation Theory, April, 1982.*

[3] Beilinson A., Bernstein J., and Deligne P., Faisceaux perverse, *Astérisque* **100** (1982).

[4] Brylinski J. L., (Co)homologie d'intersection et faisceaux pervers, *Sem. Bourbaki exposé* **585**, fevrier 1982.

[5] Brylinski J. L. and Kashiwara M., Kazhdan–Lusztig Conjecture and Holonomic Systems, *Inv. Math.* **64** (1981), pp. 387–410.

[6] Deligne P., La conjecture de Weil II, *Publ. Math. IHES* **52** (1980), pp. 137–252.

[7] Kazhdan D. and Lusztig G., Representation of Coxeter Groups and Hecke Algebras, *Inv. Math.* **53** (1979), pp. 165–184.

[8] Kazhdan D. and Lusztig G., Schubert Varietes and Poincaré Duality, *Proc. Symp. Pure Math.* **36**, pp. 185–203.

[9] Lusztig G. and Vogan D., Singularities of Closures of K-Orbits on Flag Manifolds, *Inv. Math.* **71** (1983), pp. 365–380.

[10] Springer T., Quelques applications de la cohomologie d'intersection, *Sem. Bourbaki exposé* **589**, fevrior 1982.

[11] Verdier I. L., Specialisation de faisceaux et monodromie moderée, *Astérisque* **101–102** (1983), pp. 332–364.

[12] Vogan D., Irreducible Characters of Semisimple Lie Groups III. Proof of Kazhdan–Lusztig Conjecture in the Integral Case, *Inv. Math.* **71** (1983), pp. 380–418.

[13] Vogan D., *Representation of Real Algebraic Groups*, Birkhäuser, 1981.

CARDIOLOGICAL SCIENCE CENTER
MOSCOW, USSR

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

WILLIAM FULTON

Some Aspects of Positivity in Algebraic Geometry

This article discusses variations on the following theme: *Given an imbedding $M \hookrightarrow P$, and a morphism $f\colon V \to P$, suitable "positivity" of the imbedding of M in P should force the pair $(V, f^{-1}(M))$ to "look like" the pair (P, M).* Our positivity assumption is that the normal bundle to M in P be an *ample* vector bundle. If V is an n-dimensional variety, and $d = \operatorname{codim}(M, P)$, one looks for the following sorts of conclusions. If $n \geqslant d$, $f^{-1}(M)$ should be *non-empty*. If $n > d$, $f^{-1}(M)$ should be *connected*, and the induced homomorphism of fundamental groups $\pi_1(f^{-1}(M)) \to \pi_1(V)$ surjective. More generally, one hopes to compare the groups $\pi_i(V, f^{-1}(M))$ to $\pi_i(P, M)$ for $i \leqslant n-d$. When V is not complete (compact), one expects similar conclusions, provided M is replaced by a suitable neighborhood. Finally, singularities of V or positive dimensional fibres of f tend to weaken the conclusions.

We do not know a general theorem which implies all results of this sort which have been proved. Apparently some homogeneity condition is needed in addition to the ampleness of the normal bundle.

When M is a linear subspace of a projective space P, such results include classical theorems of Bertini, Lefschetz, and Zariski: we review these, together with some modern improvements, in § 1. When $M \to P$ is a diagonal imbedding of a projective space in a product of projective spaces (§ 2), we have called such results "connectedness theorems". A third situation we shall discuss (§ 3) is when P is an ample vector bundle on a variety M, and M is imbedded in P by a section.

The connectedness theorems, although easily deduced from the classical case of linear sections, have had important applications: to singularities of mappings, to branched coverings and Gauss mappings, to a solution of Zariski's problem on the fundamental group of the complement of a plane nodal curve, to varieties of small codimension, and to a simplification and strengthening of the Barth–Larsen theorem. The vector bundle case has been applied to degeneracy loci such as varieties of special divisors, and to the problem of finding which polynomials in Chern classes of an ample vector bundle are numerically positive. These applications are discussed in § 4.

We refer to [8] for details and references for work on these questions up to 1980; the present bibliography contains some more recent articles.

§ 1. The classical case

The first statements of this kind were for a *generic* linear subspace $M = \boldsymbol{P}^{m-d}$ of $P = \boldsymbol{P}^m$. If $V^n \subset \boldsymbol{P}^m$ is closed and irreducible, one knows that such M meets V transversally in $\deg(V)$ points if $n = d$. A Bertini theorem is that $V \cap M$ is irreducible if $n > d$. The classical Lefschetz theorem is that $H_i(V, V \cap M) = 0$ for $i \leqslant n-d$, if V is non-singular. If $V \subset \boldsymbol{P}^m$ is the complement of a hypersurface, so $n = m$, a theorem of Zariski asserts that $\pi_1(V \cap M) \to \pi_1(V)$ is surjective for $n > d$, and an isomorphism if $n > d+1$. Hamm and Lê generalized this to showing that $\pi_i(V, V \cap M) = 0$ for $i \leqslant n-d$.

For an *arbitrary* linear subspace $M = \boldsymbol{P}^{n-d}$ of $P = \boldsymbol{P}^m$, and a morphism $f\colon V \to P$, these have been strengthened as follows. Denote by M_ε a small ε-neighborhood of M in P for some Riemannian metric on P. Note that if V is complete, or if M is generic, then the inclusion of $f^{-1}(M)$ in $f^{-1}(M_\varepsilon)$ is a homotopy equivalence.

(a) *If* $\dim f(V) \geqslant d$, *then* $f^{-1}(M_\varepsilon) \neq \emptyset$.

(b) *If* $\dim f(V) > d$, *then* $f^{-1}(M_\varepsilon)$ *is connected.*

(c) *If V is locally irreducible in the complex topology, and* $\dim f(V) > d$, *then* $\pi_1(f^{-1}(M_\varepsilon))$ *surjects onto* $\pi_1(V)$.

(d) *If V is locally a complete intersection of dimension n, and f is quasi-finite (i.e., all fibres are finite), then* $\pi_i(V, f^{-1}(M_\varepsilon)) = 0$ *for* $i \leqslant n-d$.

Of course, (a) is a basic fact in projective algebraic geometry. Zariski's connectedness theorem is used to deduce (b) from Bertini's theorem. Then (c), for the algebraic fundamental group, follows by applying (b) to finite coverings of V. The general case of (c) for the topological fundamental group is due to Deligne [2]. When V is a closed non-singular subvariety of $\boldsymbol{P}^m$, (d) was proved by Andreotti and Frankel. When V is the complement of a hypersurface in $\boldsymbol{P}^m$, theorems like (c) and (d) were proved by Hamm and Lê. Deligne [2] stated and proved (d) for V an arbitrary Zariski open subvariety of $\boldsymbol{P}^m$, and conjectured (d) for V non-singular. Goresky and MacPherson [12] have announced generalizations of (d), allowing fibres of positive dimension (as conjectured by Deligne) and arbitrary singularities on V.

§ 2. Connectedness theorems

In this section $M = \Delta = \boldsymbol{P}^m$ is imbedded by the diagonal in the product $P = \boldsymbol{P}^m \times \ldots \times \boldsymbol{P}^m$ of r copies of $\boldsymbol{P}^m$, so $d = (r-1)m$. As in the previous section, Δ_ε denotes a small ε-neighborhood of Δ in P; if V is complete, the inclusion of $f^{-1}(\Delta)$ in $f^{-1}(\Delta_\varepsilon)$ is a homotopy equivalence.

(a) *If V is complete, and* $\dim f(V) \geqslant d$, *then* $f^{-1}(\Delta)$ *is not empty.*

(b) *If V is complete, and* $\dim f(V) > d$, *then* $f^{-1}(\Delta)$ *is connected.*

(c) *If V is locally irreducible, and* $\dim f(V) > d$, *then* $f^{-1}(\Delta_\varepsilon)$ *is connected, and*

$$\pi_1\big(f^{-1}(\Delta_\varepsilon)\big) \to \pi_1(V)$$

is surjective.

(d) *Let V be locally a complete intersection of pure dimension n, $f\colon V \to \boldsymbol{P}^m \times \boldsymbol{P}^m$ quasi-finite; and set $d = m$.*

(i) *If $i \leqslant n-d$, $i \neq 2$, then* $\pi_i\big(V, f^{-1}(\Delta_\varepsilon)\big) = 0$.

(ii) *If $n-d \geqslant 2$, one has an exact sequence*

$$\pi_2\big(f^{-1}(\Delta_\varepsilon)\big) \to \pi_2'(V) \to \boldsymbol{Z} \to \pi_1\big(f^{-1}(\Delta_\varepsilon)\big) \to \pi_1(V) \to 0.$$

Again (a) is standard, (b) was the main theorem of [6], and (c) appeared in [2]. Deligne also showed how (d) and the other results of this section for $\Delta \subset \boldsymbol{P}^m \times \ldots \times \boldsymbol{P}^m$ can be deduced from corresponding results for a linear $\boldsymbol{P}^m \subset \boldsymbol{P}^{r(m+1)-1}$. In fact, if $r = 2$ and $x_0, \ldots, x_m, y_0, \ldots, y_m$ are homogeneous coordinates on $\boldsymbol{P}^{2m+1}$, let U be the open set where some $x_i y_j \neq 0$, and let $L \subset U$ be the linear m-space where $x_i = y_i$, $0 \leqslant i \leqslant m$. There is a canonical projection of U onto $\boldsymbol{P}^m \times \boldsymbol{P}^m$, taking $[x, y]$ to $[x] \times [y]$, which makes U a $\boldsymbol{C}^*$-bundle over $\boldsymbol{P}^m \times \boldsymbol{P}^m$ and which maps L isomorphically onto Δ. If $f\colon V \to P$ with V complete and $P = \boldsymbol{P}^m \times \boldsymbol{P}^m$, one has an induced map $\tilde f$ from $\tilde V = V \times_P U$ to $U \subset \boldsymbol{P}^{2m+1}$ and $\tilde f^{-1}(L) \cong f^{-1}(\Delta)$. One knows about $\pi_i\big(\tilde V, \tilde f^{-1}(L)\big)$ by the case of linear subspaces (§ 1), and one can compare $\pi_i(\tilde V)$ to $\pi_i(V)$ since $\tilde V$ is a $\boldsymbol{C}^*$-bundle over V.

Generalizations to flag manifolds and other homogeneous spaces have been given by Hansen, Sommese, Goldstein [11], and Faltings [4].

§ 3. Vector bundles

In this section M is an arbitrary projective variety, and $P = E$ is a vector bundle of rank d on M.

(a) *Let* $s: M \to E$ *be a section,* V *an* n*-dimensional subvariety of* E*. If* E *is ample, and* $n \geqslant d$*, then* $s^{-1}(V)$ *is not empty.*

In fact, with some intersection theory, (a) may be strengthened as follows [10]. Assume for simplicity that $n = d$. Then one may construct the *intersection product* of V by $s(M)$ on E, obtaining a well-defined rational equivalence class of zero-cycles $s^*[V]$ on M: $s^*[V] = \sum n_P[P]$ is characterized by the fact that $\sum n_P[E(P)]$ is rationally equivalent to $[V]$ on E. Note that $s^*[V]$ is independent of s:

(a′) *If* E *is ample, then* $\deg s^*[V] > 0$.

If $n > d$, we conjecture that $s^{-1}(V)$ is connected. One known case is where $E = \mathrm{Hom}(A, B) = A^\vee \otimes B$, for vector bundles A and B on M, and $V \subset E$ is the variety of maps of rank $\leqslant k$. Here $k \leqslant \min(a, b)$, where $a = \mathrm{rank} A$, $b = \mathrm{rank} B$. Then V is a subvariety of E of codimension $(a-k)(b-k)$. Specifying a homomorphism $\sigma: A \to B$ of vector bundles is equivalent to specifying a section s_σ of the bundle E, and $s_\sigma^{-1}(V) = D_k(\sigma)$, where

$$D_k(\sigma) = \{x \in M \mid \mathrm{rank}\ \sigma(x) \leqslant k\}.$$

Thus from (a), if $\mathrm{Hom}(A, B)$ is ample, and $\dim(M) \geqslant (a-k)(b-k)$, then $D_k(\sigma)$ is not empty.

(b) *If* $\mathrm{Hom}(A, B)$ *is ample, and* $\dim(M) > (a-k)(b-k)$*, then* $D_k(\sigma)$ *is connected.*

For the proof of (a′) we refer to [10], and for (b) to [9]. Griffiths, Sommese, and Lazarsfeld have proved some higher homotopy analogues, cf. [8], but less is known in this situation than in that of § 1 or § 2.

§ 4. Applications

The connectedness theorem has been applied to show that certain singularities or degeneracy will necessarily occur, provided one has reason to expect them for dimension reasons, and there is an appropriate "positivity".

4.1. Let X be irreducible n-dimensional variety, $f\colon X \to \boldsymbol{P}^m$ a finite morphism. *If $m < 2n$, and f is unramified, then f is a closed imbedding.*

To prove this, consider the product mapping $f \times f\colon X \times X \to \boldsymbol{P}^m \times \boldsymbol{P}^m$. By the connectedness theorem § 2 (b), $(f \times f)^{-1}(\Delta_{\boldsymbol{P}^m})$ is connected. But if f is unramified, Δ_X is open as well as closed in $(f \times f)^{-1}(\Delta_{\boldsymbol{P}^m})$; thus there can be no pairs (x, y) with $x \neq y$ and $f(x) = f(y)$, and the conclusion follows.

This argument was discovered by J. Hansen [6]. Other applications have followed this model, applying the connectedness theorem to mappings of appropriate product varieties to products of $\boldsymbol{P}^m$.

4.2. If $f\colon X^n \to \boldsymbol{P}^n$ is a finite branched covering of degree d, Gaffney and Lazarsfeld showed that there must be some point in X at which at least $\min(d, n+1)$ sheets come together.

4.3. If $f\colon X^2 \to \boldsymbol{P}^2$ is a finite covering branched along a nodal curve D, this writer showed that any two irreducible components of $f^{-1}(D)$ must intersect. By an argument of Abhyankar's, it follows that the algebraic fundamental group of $\boldsymbol{P}^2 - D$ is Abelian. Deligne applied § 2 (c) to the maps

$$(\boldsymbol{P}^2 - D) \times \tilde{C} \to \boldsymbol{P}^2 \times \boldsymbol{P}^2,$$

where $\tilde{C}$ is the non-singular model of an irreducible component C of D, to show that $\pi_1(\boldsymbol{P}^2 - D)$ is Abelian. A local generalization has been proved by Lê and Saito [13].

M. Nori (14) has greatly extended the solution of Zariski's problem, allowing one to replace $\boldsymbol{P}^2$ be any non-singular surface X, provided that, for each component C of the nodal curve D on X, the self-intersection number C^2 is greater than twice the number of nodes of C, i.e. the normal bundle of the immersion $\tilde{C} \to X$ is ample. With this assumption, Nori proves that

$$\operatorname{Ker}\big(\pi_1(X - D) \to \pi_1(X)\big)$$

is Abelian. For this he develops a "weak Lefschetz theorem", which implies that if $f\colon U \to X$ is a tubular neighborhood of the immersion of $\tilde{C}$ in X, then the image of $\pi_1\big(U - f^{-1}(D)\big)$ in $\pi_1(X - D)$ has *finite index* in $\pi_1(X - D)$.

4.4. F. L. Zak [17], cf. [8], used the connectedness theorem to prove that a non-singular $X^n \subset \boldsymbol{P}^m$ cannot be projected isomorphically to

P^{m-1} if $3n > 2(m-2)$. This had been conjectured by Hartshorne, who pointed out three examples of varieties with $3n = 2(m-2)$ which do project: the Veronese $P^2 \hookrightarrow P^5$, the Segre $P^2 \times P^2 \hookrightarrow P^8$, and the Plücker $G_1(P^5) \hookrightarrow P^{14}$. Lazarsfeld has produced a fourth such example, an X^{16} in P^{26}, and Zak has proved the remarkable theorem that these are the *only* four examples. Faltings [5] has approached Hartshorne's conjecture from another point of view.

4.5. Zak [17] also proved that if $X^n \subset P^m$ is non-singular, and not contained in a hyperplane, and L is any linear subspace of P^m of dimension k, $n \leqslant k \leqslant m-1$, then

$$\dim\{x \in X \mid T_x X \subset L\} \leqslant k-n.$$

In particular, the Gauss map from X to $G_n(P^m)$ is *always finite.*

Smyth and Sommese [16] have showed that if X is a non-singular subvariety of an Abelian variety A with ample normal bundle, then the degree of the Gauss map is bounded by $e/\mathrm{codim}(X, A)$, where e is the absolute value of the Euler characteristic of X. Ran [15] showed that, for subvarieties of complete homogeneous varieties, generically finite Gauss maps are always finite. Ein [3] has proved the related result that for branched coverings $X^n \to P^n$, with X^n non-singular, the ramification divisor is always ample.

4.6. If $X \subset P^n$ is a purely n-dimensional local complete intersection, then

$$\pi_i(P^m, X) = 0 \quad \text{for } i \leqslant 2n-m+1.$$

This strengthening of the Barth–Larsen theorem follows by applying § 2 (d) to the imbedding of $X \times X$ in $P^m \times P^m$.

In fact, one can show [8] that any irreducible n-dimensional subvariety of P^m is simply connected if $2n > m$. Lazarsfeld has found analogues of Barth theorems for branched coverings of P^n.

4.7. The connectedness of degeneracy loci can be applied to the loci W_d^r of special divisors on a non-singular projective curve C of genus g [7]. If $J^{(d)}(C)$ is the variety of line bundles of degree d on C,

$$W_d^r = \{L \in J^{(d)}(C) \mid \dim H^0(C, L) \geqslant r+1\}.$$

This can be realized as a degeneracy locus of a vector bundle map $A \to B$ with $A^\vee \otimes B$ ample. From § 3 (a) one deduces another proof that

W_d^r is not empty provided the Brill–Noether number

$$\varrho = g-(r+1)(g-d+r)$$

is non-negative; from § 3 (b) it follows that W_d^r is connected if $\varrho > 0$. From this and the Petri conjecture, proved by Gieseker, it follows that W_d^r is *irreducible* if $\varrho > 0$ and C is a general curve of genus g. Some refinements are given in [7].

Arbarello and Cornalba [1] have used the connectivity of W_d^2 to prove Severi's conjecture that the space of plane nodal curves of degree d with genus g is irreducible, provided $\varrho = 3d-2g-6 > 0$.

4.8. A polynomial $P = P(c_1, \ldots, c_e)$ of weight n in variables $c_1, \ldots, c_e$, where c_i has degree i, can be uniquely expressed in the form

$$P = \sum a_\lambda \Delta_\lambda$$

the sum over partitions λ of n in integers no larger than e; for a partition λ: $\lambda_1 \geqslant \ldots \geqslant \lambda_m \geqslant 0$, Δ_λ is the *Schur polynomial*

$$\Delta_\lambda = \det(c_{\lambda_i+j-i})_{1\leqslant i,j\leqslant m}\,.$$

If E is an ample vector bundle of rank e on a variety X, and $P = \sum a_\lambda \Delta_\lambda$ is a non-zero polynomial of weight n with all $a_\lambda \geqslant 0$, then $P\big(c(E)\big)$ is *numerically positive*, i.e.,

$$\int_V P\big(c_1(E), \ldots, c_e(E)\big) > 0$$

or all n-dimensional subvarieties V of X.

To see that Δ_λ is positive on $V = X$, take a trivial bundle A of rank $n+e$ on X, and a flag $A_1 \subset \ldots \subset A_m \subset A$ of trivial subbundles, with $\operatorname{rank}(A_i) = e+i-\lambda_i$. Let H be the bundle $\operatorname{Hom}(A, E)$, and let

$$\Omega_\lambda = \{\varphi \in H \mid \dim \operatorname{Ker}(\varphi)\cap A_i \geqslant i,\ 1 \leqslant i \leqslant m\}.$$

If s is the zero section of H, one has the Giambelli–Kempf–Laksov determinantal formula

$$s^*[\Omega_\lambda] = \Delta_\lambda\big(c(E)\big)\cap[X].$$

By § 3 (a′), $\deg \Delta_\lambda\big(c(E)\big)\cap[X] = \deg s^*[\Omega_\lambda] > 0$. For details, and a proof that these are the only polynomials that are numerically positive on ample vector bundles, we refer to [10].

References

[1] Arbarello E. and Cornalba M., *A Few Remarks About the Variety Irreducible Plane Curves of Given Degree and Genus*, preprint, 1982.

[2] Deligne P., Le groupe fondamental du complément d'une courbe plane n'ayant que des points doubles ordinaires est abélien, *Sém. Bourbaki 543*, Springer Lecture Notes **842** (1981), pp. 1–10.

[3] Ein L., The Ramification Divisors for Branched Coverings of P_k^n, *Math. Ann.* **261** (1982), pp. 483–485.

[4] Faltings G., Formale Geometrie und homogene Räume, *Inventiones Math.* **64** (1981), pp. 123–165.

[5] Faltings G., Verschwindungssätze und Untermannigfaltigkeiten kleiner Kodimension des komplex-projektiven Raumes, *J. Reine Angew. Math.* **326** (1981), pp. 136–151.

[6] Fulton W. and Hansen J., A Connectedness Theorem for Projective Varieties, with Applications to Intersections and Singularities of Mappings, *Ann. of Math.* **110** (1979), pp. 159–166.

[7] Fulton W., Harris J., and Lazarsfeld R., *Excess Linear Series on an Algebraic Curve*, preprint, 1983.

[8] Fulton W. and Lazarsfeld R., *Connectivity and Its Applications in Algebraic Geometry*, Springer Lecture Notes **862** (1981), pp. 26–92.

[9] Fulton W. and Lazarsfeld R., On the Connectedness of Degeneracy Loci and Special Divisors, *Acta. Math.* **146** (1981), pp. 271–283.

[10] Fulton W. and Lazarsfeld R., Positive Polynomials for Ample Vector Bundles, to appear in *Ann. of Math.*.

[11] Goldstein N., Ampleness in Complex Homogeneous Spaces and a Second Lefschetz Theorem, *Pacific J. Math.* **106** (1983), pp. 271–291.

[12] Goresky M. and MacPherson R., Stratified Morse Theory, *Proc. Symp. Pure Math. AMS* **40**, Part I (1983), pp. 517–533.

[13] Lê Dũng Tráng and Saito K., The Local π_1 of the Complement of a Hypersurface with Normal Crossings in Codimension 1 is Abelian, *RIMS-350*, Kyoto University, 1981.

[14] Nori M., *Zariski's Conjecture and Related Problems*, preprint, 1982.

[15] Ran Z., *Generically-Finite Gauss Maps are Finite*, preprint, 1982.

[16] Smyth B. and Sommese A. J., *On the Degree of the Gauss Mapping of a Submanifold of an Abelian Variety*, preprint, 1983.

[17] Zak F. L., Projections of Algebraic Varieties (in Russian), *Mat. Sb.* (*N.S.*) **116** (1981), pp. 593–602.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

JOE HARRIS

Recent Work on $\mathcal{M}_g$

In this article I will discuss some of the recent progress and still-open questions regarding the geometry of the moduli space $\mathcal{M}_g$ of smooth algebraic curves of genus g.

Defined a priori to be simply the set of isomorphism classes of compact Riemann surfaces of genus g, $\mathcal{M}_g$ may in fact be given the structure of an algebraic variety, compatible with the basic requirement that

(*) *For any map π: $X \to B$ of analytic spaces whose fibers X_b are all compact Riemann surfaces, the induced map $B \to \mathcal{M}_g$ (sending b to $[X_b]$) is holomorphic.*

19th century geometers seem to have taken the existence of such a structure for granted, and gone ahead to establish such basic facts as that $\mathcal{M}_g$ is irreducible of dimension $3g-3$. More recently, two developments have greatly advanced our understanding of $\mathcal{M}_g$. These are:

(i) the development of deformation theory, of both smooth and singular curves, which has helped us to understand better the local structure of $\mathcal{M}_g$; and

(ii) the discovery/construction of a compactification of $\mathcal{M}_g$ to a projective variety $\overline{\mathcal{M}}_g$, in which every point actually corresponds to a unique isomorphism class of curve, and which still satisfies the analogue of the condition (*) (cf. [2], [13]). This remarkable result has made it possible to apply to the study of $\mathcal{M}_g$ the many techniques developed for dealing with projective varieties.

Together, these developments have led to many of the results mentioned below (except for those of § 1).

§ 1. The topology of $\mathcal{M}_g$

There has been a great deal of progress recently in understanding the topology of $\mathcal{M}_g$, due to Harer [4] and Miller [11]. The basic approach has been to realize $\mathcal{M}_g$ as the quotient of a ball in $\boldsymbol{C}^{3g-3}$ by the mapping class group — the group of isotopy classes of homeomorphisms of the underlying topological manifold—and thus translate questions about the homology of $\mathcal{M}_g$ into questions about the topology of 2-manifolds. Roughly, the results are along 3 lines:

(i) Harer's direct computations of $H^i(\mathcal{M}_g, \boldsymbol{Z})$ mod torsion for $i = 1$ and 2 have shown that $H^1(\mathcal{M}_g, \boldsymbol{Z}) = 0$ while $H^2(\mathcal{M}_g, \boldsymbol{Z}) = \boldsymbol{Z}$ for $g \geqslant 3$ (and hence in particular that the group of divisor classes on $\mathcal{M}_g$ is $\boldsymbol{Z}$).

(ii) Stability theorems: it is found that there exists a natural indentification (mod torsion) of $H^k(\mathcal{M}_g)$ with $H^k(\mathcal{M}_{g+1})$ for $g \geqslant 3k-1$. Indeed, Miller has shown that the ring $H^*(\mathcal{M}) =: \lim\limits_{g} H^*(\mathcal{M}_g, \boldsymbol{Q})$ constructed in this way is a finite tensor product of polynomial algebras on generators of even degree and exterior algebras on generators of odd degree. Mumford in [14] suggests that $H^*(\mathcal{M})$ may actually be generated by the classes $\varkappa_i = \pi_*\big(c_1(\omega)^{i+1}\big) \in H^{2i}(\mathcal{M})$, where $\mathcal{C}_g$ is the moduli of pointed curves (C, p) (the "universal curve" over $\mathcal{M}_g$), $\pi: \mathcal{C}_g \to \mathcal{M}_g$ the natural map, and $\omega = \omega_{\mathcal{C}_g/\mathcal{M}_g}$ the relative dualizing sheaf.

(iii) Bounding the cohomological dimension of $\mathcal{M}_g$: Harer has shown that $H_k(\mathcal{M}_g) = 0$ for $k > 4g-3$; it is not known what is actually the highest non-zero homology group of $\mathcal{M}_g$.

§ 2. Complete subvarieties of $\mathcal{M}_g$

$\mathcal{M}_g$ is neither an affine variety nor a projective one. As one measure of where $\mathcal{M}_g$ falls between those two, one can ask, what is the largest dimension of a projective subvariety of $\mathcal{M}_g$? The state of our knowledge at present is this:

(i) For any $g \geqslant 3$, $\mathcal{M}_g$ contains many complete curves. Indeed, inasmuch as the Satake compactification $\tilde{\mathcal{M}}_g$ of $\mathcal{M}_g$ is projective and the boundary $\tilde{\mathcal{M}}_g - \mathcal{M}_g$ has codimension 2, we see that simply by intersecting $\tilde{\mathcal{M}}_g \subset \boldsymbol{P}^N$ with hypersurfaces we can find a complete curve in $\mathcal{M}_g$ passing through any given finite set of points. It should be said, however, that despite the apparent ubiquity of such curves, no one has yet written down explicitly a complete curve in, for example, $\mathcal{M}_3$.

(ii) Given any k, there exists a g such that $\mathcal{M}_g$ contains complete

k-dimensional varieties. One possible construction is this: starting with any curve C_0 of genus g_0, take $C_1 \xrightarrow{\pi} C_0$ an unramified double cover of C_0; then consider all double covers of C_1 ramified exactly over a divisor $\pi^{-1}(p)$, $p \in C_0$. This gives a complete curve in $\mathcal{M}_g$ for $g = 4g_0 - 2$; iterating the process k times gives a complete variety of dimension k in $\mathcal{M}_g$ for $g = 4^k g_0 - \frac{2}{3}(4^k - 1)$. (Observe however that all the subvarieties of $\mathcal{M}_g$ constructed in this way lie in the locus in $\mathcal{M}_g$ of curves with automorphisms. For all we know, it is quite possible that through a general point of $\mathcal{M}_g$ there passes no complete subvariety of dimension greater than one.)

(iii) There do not exist complete subvarieties of $\mathcal{M}_g$ of dimension $2g-1$ or greater. This follows immediately from Harer's result (§ 1, (iii)) quoted above.

There is, I should say, a very compelling reason (apart from natural curiosity) for wanting to know what is in fact the maximal dimension of a projective subvariety of $\mathcal{M}_g$ – and specifically, whether it is greater than or less than g. This is the observed fact that, whenever it has been possible to actually compute the dimensions, the image in $\mathcal{M}_g$ of any component of the variety $\mathcal{H}_{d,g,r}$ of curves of degree d and genus g in $\boldsymbol{P}^r$ has dimension greater than g. (The same statement could be made about the variety $\mathcal{G}_d^r$ of linear series of degree d and dimension r on curves of genus g.) The extremal cases here appear to be Castelnuovo curves (curves of maximal genus for their degree in $\boldsymbol{P}^r$); these have genus

$$g = \binom{m}{2}(r-1) + m\varepsilon,$$

where $d = m(r-1) + \varepsilon + 1$, $0 \leqslant \varepsilon \leqslant r-1$, and their locus in $\tilde{\mathcal{M}}_g$ has dimension

$$\binom{m+1}{2}(r-1) + (m+2)(\varepsilon+2) - r - 6 = g + (m-1)(r+1) + 2\varepsilon - 1$$

(cf. [5]). Thus, to say that $\mathcal{M}_g$ contains no complete subvarieties of dimension g or greater should imply that *the closure in* $\overline{\mathcal{M}}_g$ *of the image of every component of* $\mathcal{G}_d^r$ *meets* $\overline{\mathcal{M}}_g - \mathcal{M}_g$. If we have, as indeed we hope to, some description of a compactification of $\mathcal{G}_d^r$ (cf. [3], [9]) it might then be possible to approach the problem of estimating the dimension (and deciding the irreducibility) of $\mathcal{G}_d^r$ and $\mathcal{H}_{d,g,r}$ via a specialization to singular curves.

§ 3. Stratifications

One approach to understanding the geometry of $\mathcal{M}_g$ and $\mathcal{C}_g$ is via stratifications, expressing $\mathcal{C}_g$ as a union of locally closed subvarieties of simpler structure. The classic example of this is the Arbarello stratification, obtained by considering the loci

$$\mathscr{C}_g^{(k)} = \left\{(C, p) : \begin{matrix} h^0(C, \mathcal{O}(kp)) = 2 \\ h^0(C, \mathcal{O}((k-1)p)) = 1 \end{matrix}\right\} \cdots$$

that is, the locus of pairs (C, p) such that C is expressible as a k-sheeted cover of $\boldsymbol{P}^1$ with p a point of total ramification. Arbarello in [1] shows that $\mathscr{C}_g^{(k)}$ is irreducible of dimension $2g-3+k$ and contained in the closure in $\mathscr{C}_g$ of $\mathscr{C}_g^{(k+1)}$, so that the $\mathscr{C}_g^{(k)}$ give a simple and potentially very useful stratification of $\mathscr{C}_g$.

The problem here is that not enough is known about the strata. Arbarello in his thesis suggested that in fact $\mathscr{C}_g^{(k)}$ might be affine, but this appears to be not the case in general.[1] Still open, and very much of interest, is the weaker conjecture that $\mathscr{C}_g^{(k)}$ is pseudo-ample in the closure of $\mathscr{C}_g^{(k+1)}$ in $\mathscr{C}_g$, i.e. that $\mathscr{C}_g^{(k)}$ contains no complete curves. To illustrate the use of stratifications, observe that if this could be shown, it would immediately follow that $\mathscr{C}_g$ contained no complete g-dimensional subvarieties (and hence $\mathscr{M}_g$ no complete $(g-1)$-dimensional ones): such a subvariety would have to meet $\mathscr{C}_g^{(k)}$ in a complete $(k-1)$-dimensional variety; but $\mathscr{C}_g^{(2)}$, a finite covering of the hyperelliptic locus, is affine.

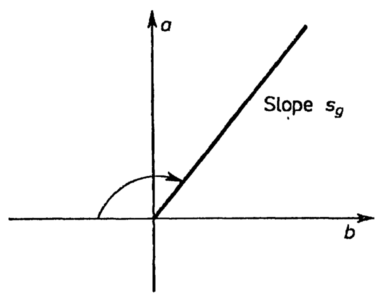
Another possible stratification is a refinement of the Arbarello stratification: for any semigroup $H \subset \boldsymbol{Z}^+$ of index g, we let $\mathscr{C}_g^H \subset \mathscr{C}_g$ be the locus of pairs (C, p) such that the Weierstrass semigroup $H_{C,p}$ of orders of poles of meromorphic functions on C holomorphic on $C-p$ is H. Here it *is* possible to show, for example, that the strata $\mathscr{C}_g^H$ contain no complete curves ([7]), and one may even conjecture that they are affine. The problem is with the configuration of the strata themselves: no one knows $\dim \mathscr{C}_g^H$ in general, or for which pairs H, H' we have $\mathscr{C}_g^{H'} \subset \overline{\mathscr{C}_g^H}$, or for that matter for which $H \mathscr{C}_g^H$ is nonempty.

§ 4. Divisors on $\overline{\mathscr{M}}_g$

Thanks to the result of Harer on $H^2(\mathscr{M}_g)$, we now know that the group $\operatorname{Pic}(\overline{\mathscr{M}}_g)$ of divisors on $\overline{\mathscr{M}}_g$ is generated freely by the class λ of the Hodge bundle $\pi_* \omega_{\overline{\mathscr{C}}_g/\overline{\mathscr{M}}_g}$ and the class δ_i of the various boundary components Δ_i (cf. [13]). Many questions about the geometry of divisors on $\overline{\mathscr{M}}_g$ remain,

[1] By examining various curves in the closure of $\mathscr{C}_g^{(k)}$ in $\overline{\mathscr{C}}_g$, one can see that no linear combination of the components of the divisor $\overline{\mathscr{C}_g^{(k)}} - \mathscr{C}_g^{(k)}$ can be ample.

of course; perhaps the biggest of which is to describe the cone of effective divisors in $\overline{\mathcal{M}}_g$. Here for simplicity, let us restrict our attention to the subgroup $\{a\lambda - b\delta\}$ of $\mathrm{Pic}(\overline{\mathcal{M}}_g)\otimes\mathbf{Q}$ generated by λ and $\delta = \sum\delta_i$. We know then that the cone (the saturation of the semigroup) of effective divisors looks like



but we do not know the value of the slope s_g. What is known is that

(i) $s_g \leqslant 6+\dfrac{12}{g+1}$ for g odd (cf. [9]), and $s_g \leqslant \dfrac{6g^2+22g-84}{g^2-6}$ for g even (cf. [6]),

(ii) $s_g \geqslant 6\frac{1}{2}$ for $g \leqslant 12$.

Of course, part of the significance of this question stems from the determination of the canonical class of $\overline{\mathcal{M}}_g$:

$$K_{\overline{\mathcal{M}}_g} \equiv 13\lambda - 12\delta$$

which leads to the equivalences $s_g \leqslant 6\frac{1}{2} \Leftrightarrow$ the Kodaira number $K_{\overline{\mathcal{M}}_g} \geqslant 0$ and $s_g < 6\frac{1}{2} \Leftrightarrow$ the Kodaira number $K_{\overline{\mathcal{M}}_g} = 3g-3$ (and which leads to (ii) above). One question of interest in particular is about the limit

$$s = \liminf_{g\to\infty} s_g.$$

We know that $6 \geqslant s \geqslant 0$, but no more; it would be nice to know, for example, whether $s = 0$ or $s > 0$. One reason for asking this is that $s > 0$ would imply that certain automorphic forms for the symplectic group $\mathrm{Sp}(2g, \mathbf{Z})$ acting on the Seigel upper half plane, viewed as forms on the moduli space $\mathcal{A}_g$ of Abelian varieties, vanished identically on the locus $\mathcal{M}_g \subset \mathcal{A}_g$, giving new equations for $\mathcal{M}_g$.

One approach to the problem of finding s_g would be via curves in $\overline{\mathcal{M}}_g$. Explicitly, if C is any curve in $\overline{\mathcal{M}}_g$ whose deformations fill up an open subset of $\overline{\mathscr{M}}_g$, then it follows that

$$s_g \geqslant \frac{\deg_C \delta}{\deg_C \lambda}.$$

This was worked out in one case in [8], where the curve C was constructed by taking a general 1-parameter family $\{B_\lambda\}$ of b-tuples of points on P^1, and letting C be the locus of all k-sheeted covers of P^1 simply branched over B_λ for some λ. It was found that

$$\frac{\deg_C}{\deg_C} \approx \frac{72(b-1)}{(b-1)(2k+5)-\frac{9}{2}k(k-1)}.$$

Since C will fill out an open subset of $\overline{\mathscr{M}}_g$ whenever $g \leqslant 2k-2$, it followed that $s_g \geqslant s_g^0$, where

$$s_g^0 = \frac{576}{5g} + o\left(\frac{1}{g}\right).$$

§ 5. Kodaira dimension of $\mathscr{M}_g$

The question of for which g the moduli space $\overline{\mathscr{M}}_g$ is rational, stably rational, unirational, uniruled and/or of negative Kodaira number has received a good deal of attention of late. This is, to the best of my knowledge, the state of affairs.

$\mathscr{M}_g$ was known classically to be unirational for $g \leqslant 10$. In addition, it was recently shown to be unirational in case $g = 12$ by Sernesi [15] and to be uniruled if $g = 11$ by Mori [12]. As regards rationality, it was classically known that $\mathscr{M}_2$ is rational and $\mathscr{M}_3$, $\mathscr{M}_4$ and $\mathscr{M}_5$ stably rational; recently Kollar and Schreyer have shown that $\mathscr{M}_6$ is stably rational as well [10]. All of these results follow from an examination of a projective model of a general curve, either as a plane curve ([15]), a canonical curve ([10]) or a hyperplane section of a $K-3$ surface ([12]). It is not at all clear how much further one can go along these lines.

In the other direction, the fact that $\mathscr{M}_g$ is not unirational was established for odd $g \geqslant 23$ in [9] and for even $g \geqslant 40$ in [6]; David Eisenbud and I hope to apply similar techniques to show ultimately that $\mathscr{M}_g$ of general type for all $g \geqslant 24$. The technique used in obtaining all these results has nothing to do per se with projective models; rather, as indicated above, it consists of exhibiting effective pluricanonical divisors. Thus, for example, for $g = 2k-1$ the closure $\overline{D_k}$ in $\overline{\mathscr{M}}_g$ of the divisor

$D_k \subset \mathcal{M}_g$ of curves which admit a g_k^1 is examined in [6], and shown to have class

$$\bar{D}_k \sim a\lambda - b_0\delta_0 - b_1\delta_1 - \ldots,$$

where

$$\frac{a}{b_0} = 6 + \frac{6}{k} = 6 + \frac{12}{g+1}.$$

This seems to be the best that one can do: in general, whenever

$$\varrho =: g-(4+1)(g-d+r) = -1$$

one would expect that the locus of curves possessing a g_d^r would form a divisor in $\mathcal{M}_g$, and I believe that the class of the closure of this divisor in $\bar{\mathcal{M}}_g$ will have "slope" $\dfrac{a}{b_0} = 6 + \dfrac{12}{g+1}$ just as in the case $r = 1$ above. Of all the divisors in $\bar{\mathcal{M}}_g$ whose classes have been determined this is the lowest slope; and I am inclined to think that it is indeed the lowest among divisors characterized by a natural geometric property of the curves represented. If so, it would seem that this approach to determining the Kodaira dimension of $\bar{\mathcal{M}}_g$ cannot be applied for $g < 23$.

In sum, then, it appears that the disposition of $\bar{\mathcal{M}}_g$ for g between 13 and 22 remains very much a mystery.

Note. Since the preparation of this manuscript, S. Diaz of the University of Pennsylvania has shown that $\mathcal{M}_g$ contains no complete $(g-1)$-dimensional subvarieties, answering the central question of § 2 above. Also M. Chang and Z. Ran have shown that $\mathcal{M}_g$ is unilateral for $g = 11$ and 13.

References

[1] Arbarello E., Weierstrass Points and Moduli of Curves, *Comp. Math.* **29** (1974), pp. 325–342.

[2] Deligne P. and Mumford D., The Irreducibility of the Space of Curves of Given Genus, *Publ. IHES* **36** (1969), p. 75.

[3] Eisenbud D. and Harris J., A Simpler Proof of the Gieseker–Petri Theorem on Special Divisors, to appear in *Invent. Math.*

[4] Harer J., The Second Homology Group of the Mapping Class Group of an Orientable Surface, *Invent. Math.* **72** (1983), pp. 221–239.

[5] Harris J., *Curves in Projective Space*, Les Presses de l'Université de Montréal, 1982.

[6] Harris J., *On the Kodaira Dimension of the Moduli Space of Curves II: the Even-Genus Case*, preprint.

[7] Harris J., *Complete Families of Weierstrass Points*, preprint.

[8] Harris J. and Morrison I., in preparation.
[9] Harris J. and Mumford D., On the Kodaira Dimension of the Moduli Space of Curves, *Invent. Math.* **67** (1982), pp. 23–86.
[10] Kollar J. and Schreyer F., The Moduli of Curves is Stably Rational for $g \leqslant 6$, to appear in *Duke Math. J.*
[11] Miller E., *The Homology of the Mapping Class Group of Surfaces*, in preparation.
[12] Mori S., preprint.
[13] Mumford D., Stability of Projective Varieties, *L'Ens. Math.* **24** (1977), pp. 39–210.
[14] Mumford D., *Toward an Enumerative Geometry of the Moduli Space of Curves*, preprint.
[15] Sernesi E., Unirationality of the Variety of Moduli of Curves of Genus 12, *Ann. Scuola Norm. Sup. Pisa* **8** (1981), pp. 405–439.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

SHIGERU IITAKA

Birational Geometry of Algebraic Varieties

This is a continuation of articles read by B. Moishezohn (1974, Nice) and K. Ueno (1978, Helsinki). Our purpose is to present a perspective view on the recent progress of the birational classification theory of algebraic varieties.

For simplicity, we restrict ourselves to algebraic varieties defined over the field of complex numbers.

§ 1. Preliminary

Needless to say, the prototype of classification theory of varieties is the classical classification theory of algebraic surfaces by the Italian school, enriched by Zariski, Kodaira and others.

Let me start by recalling the basic notions. Given a variety V, we have a non-singular model by Hironaka; this implies that there exist a non-singular variety V_1 and a proper birational map $\mu: V_1 \to V$. By using this, we define the following basic birational invariants:

$$P_m(V) := \dim H^0(V_1, \omega^m) \quad \text{for any } m > 0,$$

and

$$q(V) := \dim H^0(V_1, \Omega^1).$$

Here, Ω^1 denotes the sheaf of regular differential 1-forms and ω^m the m-times tensor product of the sheaf of regular n-forms, n being $\dim V$. Then the *Kodaira dimension* of V is defined to be the non-negative integer $\varkappa$ satisfying the following estimate: whenever $P_{m_0}(V) > 0$ for some positive integer m_0,

$$\alpha m^{\varkappa} \geqslant P_{mm_0}(V) \geqslant \beta m^{\varkappa}.$$

Here α and β are positive numbers and m is sufficiently large. If $P_m(V) = 0$ for all m, the Kodaira dimension is defined to be $-\infty$.

The Kodaira dimension is a birational invariant which takes one of the values $-\infty, 0, 1, \ldots, n$. This is the most basic birational invariant in this theory. The fundamental structures of V are revealed by studying their Kodaira dimension.

§ 2. Fundamental results

Let f: $V \to W$ be an *algebraic fiber space*, i.e., f is a proper and surjective morphism with connected fibers. We denote the dimensions of V and W by n and m, respectively.

The following facts are easily proved (cf. [5]):

EASY ADDITION THEOREM. *For an algebraic fiber space in the above sense, we have*

$$\varkappa(V) \leqslant \varkappa(F_w) + \dim W.$$

Here, F_w is a general fiber $= f^{-1}(w)$, w being a general point of W.

Note that by the definition of Kodaira dimension if $\varkappa(F_w) = -\infty$, then $\varkappa(V) = -\infty$.

COVERING LEMMA. *If $V_1 \to V$ is an étale covering, then $\varkappa(V_1) = \varkappa(V)$.*

CANONICAL FIBERING THEOREM. *If $\varkappa(V) \geqslant 0$, then there exist a non-singular variety V^*, a birational map μ: $V^* \to V$ and a fiber space f: $V^* \to W$ such that* (1) $\dim W = \varkappa(V)$, (2) $\varkappa(F_w) = 0$ *for general points w of W. Moreover, such an algebraic fiber space f is unique up to birational equivalence.*

Thanks to the canonical fibering theorem, the study of varieties V is reduced to that of V with $\varkappa(V) = -\infty$ or 0 or n, n being $\dim V$.

Unfortunately, the following fundamental conjecture has not yet been solved completely.:

CONJECTURE. $C_{n,m}$: $\varkappa(V) \geqslant \varkappa(F_w) + \varkappa(W)$.

C_n means $C_{n,m}$ for any $m > 0$.

In the following cases, the conjecture is verified:

(1) W is a variety of general type, i.e., $\varkappa(W) = \dim W$ (by Kawamata [9] and Viehweg [17]).

(2) General fibers have trivial pluri-canonical divisors (by Kawamata [11], [12]).

(3) General fibers are curves or surfaces (by Viehweg and Kawamata, see [11], [12]).

Varieties V with $\varkappa(V) = -\infty$. In the case of surfaces, such V are ruled surfaces (Enriques). If $n = 3$, $\varkappa(V) = -\infty$ and $q(V) > 0$, then there exists a (possibly ramified) covering h: $V_1 \to V$ such that V_1 is ruled. This was proved by applying C_3 (Viehweg [16]). In general if a variety V has a covering V_1 that is ruled, V is said to be a ***uniruled variety*** or a ***quasiruled variety***. The collection of polarized Kaehler varieties which are not uniruled has nice algebraic space structures (A. Fujiki [3]). Recently a variety whose anticanonical divisor is ample has been proved to be uniruled. This result follows from Mori's theory of cones of 1-cycles (see Mori's article in this volume).

Varieties V with $\kappa(V) = 0$. Then the Albanese map α: $V \to \mathrm{Alb}(V)$ is a fiber space (Kawamata [9]), i.e., α is surjective and has connected fibers. In particular, if $n = q(V)$ $(= \dim \mathrm{Alb}(V))$, then α is a birational morphism. This follows from C_n and it gives a birational characterization of abelian varieties.

A general theory of varieties V with $\kappa(V) = 0$ and $q(V) = 0$ has not yet been established. For surfaces, such V are birationally equivalent to K3 surfaces or Enriques surfaces. When V satisfies $\kappa(V) = 0$ and $0 < q(V) < n$, the Albanese fiber space $V \to \mathrm{Alb}(V)$ seems to be birationally equivalent to a fiber bundle whose general fibers are varieties F with $\kappa(F) = 0$. For 3-folds this is verified by Viehweg [16].

Varieties V with $\kappa(V) = n$. It has not yet been established that the graded ring $R(V) = \bigoplus_{i=0} H^0(V, \omega^i)$ is finitely generated. When $n = 3$ and a canonical divisor K_V is numerically semipositive (or numerically effective), i.e., $K_V \cdot C \geqslant 0$ for any curve C, this was proved by Kawamata [13] and X. Benveniste [1]. They proved a stronger result to the effect that K_V is semiample in this case, i.e., there is a pluricanonical system $|mK_V|$ which has no base points at all for some $m > 0$. This is closely related to the construction of good minimal models for threefolds.

§ 3. Birational geometry

Birational geometry is the study of birational equivalence classes of varieties. Since any variety is birationally equivalent to a complete non-singular one, it may be enough to study non-singular complete varieties. But when one investigates the structure of non-complete varieties, it is occasionally helpful to use ***strictly rational maps*** and ***proper birational maps***.

Let us recall their definitions.

Definition. A rational map $f\colon V \to W$ is said be a *strictly rational map* if the projection from the graph to V is a proper birational morphism. Further, a birational map is said to be a *proper birational map* if both f and f^{-1} are strictly rational maps.

Any rational map into a complete variety is strictly rational. If a variety V is normal and $f\colon V \to W$ is a strictly rational map, then f is defined outside a closed subset F with $\operatorname{codim}(F) \geqslant 2$. In particular, if W is affine and V is normal, every strictly rational map is a morphism.

When one uses strictly rational maps and studies proper birational equivalence classes, one has proper birational geometry. More generally, considering variations of "rational maps", "birational equivalence classes" and "non-singular models", we can develop various kinds of birational geometries.

An outline of such a birational geometry is given below:

(1) Fix a subset $\mathscr{M}$ of the set of varieties and define rational maps and birational equivalence classes in $\mathscr{M}$.

(2) Introduce non-singular objects in $\mathscr{M}$ and prove the existence of non-singular models for any object in $\mathscr{M}$.

(3) Using non-singular models, define regular forms and introduce plurigenera and Kodaira dimension for any object in $\mathscr{M}$.

(4) Obtain a rough classification of objects in $\mathscr{M}$ by means of the Kodaira dimension.

(5) Study particular structures of objects with the Kodaira dimension $-\infty$ or 0 or n, n being the dimension.

(6) Introduce the notion of minimal models and study further...

In any case the classical classification theory of algebraic surfaces is a good guide. Recall that the classification of higher dimensional varieties has been developed in this way.

(I) Let $\mathscr{M}$ be a collection of not necessarily complete varieties. Then strictly rational maps and proper birational maps are used as "rational maps" and "birational maps". For any variety V there exists a non-singular variety Y and a proper birational morphism $\mu\colon Y \to V$. It may be natural to take Y as a non-singular model of V. However, there exists a non-singular completion $\overline{Y}$ with smooth boundary D, which implies that $Y = \overline{Y} - D$ and D has only simple normal crossings. It may be better to consider the triple $(Y, \overline{Y}, D)$ as a nonsingular model of V. In this case, "regular forms" of V are logarithmic forms on $\overline{Y}$ with logarithmic poles along D. Strictly speaking, let $\Omega(\log D)$ denote the sheaf of germs of logarithmic 1-forms on $\overline{Y}$ with logarithmic poles along D. For a polynomial representation ϱ of GL(n), n being $\dim V$, $\Gamma(\bar{Y}, \Omega^{\varrho})$ is the space of logarithmic forms associated with ϱ. Indeed, these vector spaces and their dimensions are invariants under any proper birational map. In particular, the Kodaira dimension of V is defined to be $\kappa(K+D, \bar{Y})$, where K is a canonical divisor on $\bar{Y}$. This is denoted by $\bar{\kappa}(V)$. Replacing κ by $\bar{\kappa}$ and complete varieties by not necessarily complete ones, we have similar canonical fibering, called *logarithmic canonical fibering.*

The easy addition theorem, the covering lemma and the existence of canonical fibering are formulated for the logarithmic Kodaira dimension and are easily proven. A logarithmic analogue of Conjecture C_n, denoted by $\bar{C}_n$, seems more interesting. $\bar{C}_2$ is completely solved and the structure theory of open surfaces, i.e., classification of surfaces in proper birational geometry, has been studied in detail ([8], [15]). Here the notion of numerical semipositivity and some variants of minimal models play indispensable roles (for example, the notion of an almost minimal model by Tsunoda). Even $\bar{C}_3$ seems to be left unsolved. However, in the case where the base variety W is of general type, i.e., $\bar{\kappa}(W) = \dim W$, $\bar{C}_n$ has been recently verified by Maehara and Matsuda. For this, the weak-positivity of the direct image sheaf of the logarithmic relative canonical sheaf is a key fact (Viehweg [18], [19], Maehara).

(II). Affine varieties are special cases of non-complete varieties, hence they can be studied in birational geometry. But in some cases, normal varieties play roles of non-singular models.

(III). Considering pairs (D, V) of normal varieties V and reduced divisors D on V, one can develop the birational geometry of these pairs. For such a pair (D, V), there exist a non-singular variety Y and a divisor B with simple normal crossings on Y and a birational map μ: $Y \to V$ such that the strict transform of B is D. Then the logarithmic Kodaira dimension $\bar{\kappa}(Y-B) = \kappa(B+K(Y), Y)$ is a birational invariant; hence it is a birational invariant of (D, V), which we denote by $\kappa[D]$, ignoring V. When $V = \boldsymbol{P}^2$, $\kappa[D] = -\infty$ if and only if D is transformed into a line by a Cremona transformation. This is almost equivalent to Max Noether's theorem on the factorization of Cremona transformations. Further, plane curves are studied in the framework of birational geometry ([6], [7]).

References

[1] Benveniste X., Sur l'anneau canonique de certaines varieties de dimension 3, *Inv. Math.* **73** (1983), pp. 157–164.

[2] Fujiki A., On the structure of compact complex manifolds in C, *Advances in Pure Mathematics* 1, Kinokuniya, North-Holland, Tokyo 1983, pp. 231–302.

[3] Fujiki A., *Coarse moduli space for polarized compact Kaehler manifolds and polarized algebraic manifolds*, to appear.

[4] Fujita T., Kaehler fiber spaces over curves, *J. Math. Soc. Japan* **30** (1978), pp. 779–794.

[5] Iitaka S., Algebraic Geometry, *G. T. M.* **76** (1981), Springer, Berlin, Heidelberg, New York.

[6] Iitaka S., *Basic structures of algebraic varieties, Part 2. Cremona transformations and logarithmic Kodaira dimension,* Advanced Studies in Pure Mathematics, vol. 1, Kinokuniya, North-Holland, Tokyo, 1982, pp. 303–316.

[7] Iitaka S., On irreducible plane curves, *Saitama Math. J.* **1** (1983), pp. 47–63.

[8] Kawamata Y., *On the classification of non-complete algebraic surfaces, Algebraic Geometry,* Proceedings, Copenhagen 1978, Lecture Notes in Math. No. **732**, Springer, 1978, pp. 215–232.

[9] Kawamata Y., Characterization of abelian varieties, *Compositio Math.* **43** (1981) pp. 253–276.

[10] Kawamata Y., Kodaira dimension of algebraic fiber spaces over curves, *Invent. Math.* **66** (1981), pp. 57–72.

[11] Kawamata Y., Kodaira dimension of certain fiber spaces, *J. Fac. Sci. Univ. of Tokyo,* IA, **30** (1983), pp. 1–24.

[12] Kawamata Y., *Hodge theory and Kodaira dimension,* Advanced Studies in Pure Mathematics, vol. 1, Kinokuniya, North-Holland, Tokyo, 1983, pp. 317–327.

[13] Kawamata Y., *On the finiteness of generators of a pluricanonical ring for a 3-fold of general type,* preprint.

[14] Maehara K., A finiteness property of varieties of genera 1 type, *Math. Ann.* (1983), pp. 101–123.

[15] Miyanishi M., *Theory of non-complete algebraic surfaces,* Lecture Notes in Math. **857**, Springer 1981.

[16] Viehweg E., Klassifikationstheorie algebraischer Varietäten der Dimension drei, *Compositio Math.* **41** (1981), pp. 361–400.

[17] Viehweg E., Die Additivität der Kodaira Dimension für projektive Faserraume über Varietäten des allgemeinen Typs, *J. f. reine u. angew. Math.* **330** (1982), pp. 132–142.

[18] Viehweg E., *Weak positivity and the addivity of the Kodaira dimension for certain fiber spaces,* Advanced studies in pure mathematics 1, 1983, pp. 329–353.

[19] Viehweg E., *Weak positivity and the additivity of the Kodaira dimension, II: The local Torelli map, Classification of algebraic and analytic manifolds,* Katata Symposium, Proceedings 1982, (1984), pp. 567–589.

DEPARTMENT OF MATHEMATICS,
UNIVERSITY OF TOKYO,
HONGO, BUNKYO, TOKYO, 113
JAPAN

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

V. A. ISKOVSKIH

Algebraic Threefolds with Special Regard to the Problem of Rationality

In this report we discuss the same problems as those discussed in the well-known book by Roth [36], and therefore the report has the same title (see also Fano [11]).

This domain has shown a significant progress in the last ten years, connected with the discovery of new tools and the development of the old ones aimed at the proofs of the non-rationality of some types of unirational algebraic threefolds (see [1], [7], [17]). We outline here some of the up-to-date basic results bearing on the problems of rationality and unirationality of algebraic threefolds. We restrict ourselves to the following two types of threefolds:

I. Fano threefolds V with $\operatorname{Pic} V \simeq \boldsymbol{Z}$ (i.e., Fano threefolds of the first species in the terminology of [19]).

II. Conic bundles $\pi\colon V \to S$ over rational surfaces S.

The ground field k is assumed to be algebraically closed and of characteristic zero unless otherwise stated.

1. DEFINITION. A complete smooth irreducible threefold over k is a *Fano threefold* if the anticanonical sheaf K_V^{-1} of V is ample. The greatest integer $r \geqslant 1$ such that $H^r \cong K_V^{-1}$ for some invertible sheaf $H \in \operatorname{Pic} V$ is called the *index* of V. The integer $g = g(V) = -K_V^3/2+1$ is called the *genus* of V, that is, the genus of curves which are linear sections of V in the anticanonical embedding. A Fano threefold V with $\operatorname{Pic} V \simeq \boldsymbol{Z}$ is called a *threefold of the first species* (see [18], [19] or [23], [29]).

If V is a Fano threefold, then $h^1(V, O_V) = h^2(V, O_V) = h^3(V, O_V) = 0$, $\operatorname{Pic} V$ is a free abelian group of the finite rank ϱ (ϱ is the Picard number of V), and $r = 1, 2, 3$, or 4. Furthermore, if $r = 4$ then $V \simeq \boldsymbol{P}^3$, and if $r = 3$ then $V \simeq Q$ – a quadric in $\boldsymbol{P}^4$ ([18]).

We shall treat only Fano threefolds of the first species. Fano threefolds with the Picard number $\varrho \geqslant 2$ were classified up to deformation by Mori and Mukai [26] on the basis of Mori's paper [25] and are discussed in Mori's report at this congress.

The list of Fano threefolds of the first species is given in [19] and [20]. However, there is one threefold missing in [19], see A_{22} below. The author is grateful to Professor S. Mukai and Professor M. Reid for pointing out this gap. In addition, N. P. Gushel [14] has shown that, besides the variety pointed out in [19], there is only one type of Fano threefolds of the first species V with $r = 1$ and $g(V) = 6$ (see A'_{10} below). The final list of Fano threefolds of the first species and of indices $r = 1, 2$ yields the following

THEOREM 1. *Let V be a Fano threefold of the first species and let $h = b_3/2$ be half of the third Betti number of V and $d = H^3$.*

If $r = 1$, then V is one of the following:

A_2: *a double covering of $\boldsymbol{P}^3$ branched in a smooth sextic* ($d = 2$, $h = 52$, $g = 2$);

A_4: *a smooth quartic $V_4 \subset \boldsymbol{P}^4$* ($d = 4$, $h = 30$, $g = 3$);

A'_4: *a double covering of a smooth quadric $Q \subset \boldsymbol{P}^4$ branched in a surface of degree* 8 ($d = 4$, $h = 30$, $g = 3$);

A_6: *a smooth complete intersection of a quadric and a cubic in $\boldsymbol{P}^5$* ($d = 6$, $h = 20$, $g = 4$);

A_8: *a smooth complete intersection of three quadrics in $\boldsymbol{P}^6$* ($d = 8$, $h = 14$, $g = 5$);

A_{10}: *a smooth section of the grassmanian* $\mathrm{Gr}(2, 5) \subset \boldsymbol{P}^9$ *by a quadric and two hyperplanes* ($d = 10$, $h = 10$, $g = 6$, see [19], [14]);

A'_{10}: *a smooth section of the cone over B_5* (see below) *by a quadric in $\boldsymbol{P}^7$* ($d = 10$, $h = 10$, $g = 6$, see [14]);

A_{12}: *the threefold $V_{12} \subset \boldsymbol{P}^8$* ($d = 12$, $h = 5$, $g = 7$, see [19]);

A_{14}: *a section of the grassmanian* $\mathrm{Gr}(2, 6) \subset \boldsymbol{P}^{14}$ *by five hyperplanes* ($d = 14$, $h = 5$, $g = 8$, see [19], [15]);

A_{16}: *the threefold $V_{16} \subset \boldsymbol{P}^{10}$* ($d = 16$, $h = 3$, $g = 9$, see [19]);

A_{18}: *the threefold $V_{18} \subset \boldsymbol{P}^{11}$* ($d = 18$, $h = 2$, $g = 10$, see [19];

A_{22}: *the threefold $V_{22} \subset \boldsymbol{P}^{13}$* ($d = 22$, $h = 0$, $g = 12$, see [19]);

A'_{22}: *the compactification in $\boldsymbol{P}^{13}$ of the affine homogeneous variety* $\mathrm{PGL}(2)/I \subset A^{13}$ where $I \subset \mathrm{PGL}(2)$ *is the icosahedron subgroup and A^{13} is the space of binary forms of degree* 12 (here $\mathrm{PGL}(2)/I$ is identified with the orbit of a I-invariant binary form).

If $r = 2$, then V is one of the following:

B_1: *a double covering of the cone $W \subset \boldsymbol{P}^6$ over the Veronese surface*

$F^4 \subset P^5$, branched in a nonsingular divisor $D \subset W$ traced on W by a cubic in P^6 which does not pass through the vertex of the cone $(d = 1,\ h = 21,\ g = 5)$;

B_2: *a double covering* of P^3 *branched in a smooth quartic* $(d = 2,\ h = 10,\ g = 9)$;

B_3: *a smooth cubic in* P^4 $(d = 3,\ h = 5,\ g = 13)$;

B_4: *a smooth intersection of two quadrics in* P^5 $(d = 4,\ h = 2,\ g = 17)$;

B_5: *a section of the grassmanian* $\mathrm{Gr}(2, 5) \subset P^6$ *by three hyperplanes* $(d = 5,\ h = 0,\ g = 21)$.

In the proof of the theorem the results of V. V. Shokurov about the smoothness of the generic anticanonical divisor [40] and about the existence of a line are essentially used [41] (see also [35]).

2. In studying the rationality problem two methods are used: the Clemens-Griffiths method of the intermediate Jacobian [7] and the method, originated by Noether and Fano, of "untwisting" birational automorphisms ([10], [13], [17]). The Artin–Mumford method [1] does not work here, for Fano varieties V are torsion-free in $H^3(V, Z)$. Now we shall summarize, in the form of several theorems followed by brief comments and references, the results on the unirationality and rationality of Fano threefolds.

THEOREM 2. *All Fano varieties of types A_{2k} and B_l for $k \geqslant 3$ and $l \geqslant 2$ as well as varieties of types A'_4, A'_{10} and A'_{22}, are unirational, and so are some of the smooth quartics A_4.*

The unirationality (and rationality) of varieties is usually proved by means of direct geometrical constructions. Constructions used to prove the rationality of Fano varieties are on the whole classical ([12], [36]). The unirationality of some smooth quartics A_4 was proved by B. Segre [39] (see also [17]). The unirationality of A'_4, A_6 and of some of A_4 is explained in [21]. It is not known if the generic quartics and varieties of types A_2 and B_1 are unirational. This is one of the difficult and fundamental questions in the birational geometry of three-dimensional algebraic varieties.

THEOREM 3. *Varieties of types $A_{12}, A_{16}, A_{18}, A_{22}, A'_{22}, B_4,$ and B_5 are rational.*

Establishing the rationality of all these varieties (except for A'_{22}) can be found in [18], [19], [20]. The rationality of A'_{22} is also well known [50].

Now we dwell on the proof of the non-rationality of all the other types of Fano threefolds. We first note some general results proved by the degeneration principle and the Clemens–Griffiths method of the intermediate Jacobian.

THEOREM 4. *Any generic varieties of types $A_2, A_4, A_4', A_6, A_{10}, A_{10}'$, B_1, B_2 and any smooth one of types A_8, A_{14} and B_3 over C are non-rational.*

The non-rationality of an arbitrary smooth cubic has been proved as is well known by Clemens and Griffiths [7] (see also [4], [28], [46], [47]). Any of the other varieties mentioned in the theorem can be deformed or birationally transformed, as is done by Beauville [3] (see also [47]), to a net of conics whose intermediate Jacobian (which is a Prym variety) is not the Jacobian of a curve or of a product of curves. Now the condition of the local closure of the set of Jacobians among all principally polarized abelian varieties leads to the proof of the theorem.

Besides a cubic B_3, by using the Clemens–Griffiths method the non-rationality of any smooth variety of the type A_8 is proved ([3], [48]). Any such variety can be birationally represented in the form of a conic bundle over P^2 with the degeneracy curve of degree 7. Furthermore, as shown by Fano, the variety A_{14} is birationally equivalent to the cubic B_3, and hence is non-rational (see [21], [32]).

Remark 1. The problem of calculating the intermediate Jacobian for the remaining Fano varieties is very difficult. A great deal of work in this direction has been done by A. S. Tikhomirov [43], [44] and Welters [51], who evaluated the intermediate Jacobian for type B_2 and Tikhomirov [45] did this for type B_1. As far as I know, the non-rationality of any smooth varieties of these types has by now been established (but by another method, see below).[1] For the calculation of intermediate Jacobians of Fano varieties Tyurin's method [46] (see also [5]) of generalized Prym varieties and Clemens degeneration method [6] seem to be promising.

3. Now we shall consider the method of Noether–Fano of the untwisting birational automorphisms. The assumption that char $k = 0$ is not essential here, so we shall assume only that char $k \neq 2, 3$.

THEOREM 5. *For any smooth Fano variety V of type A_2, A_4 or B_1 the group of birational automorphisms* $\mathrm{Bir}\,V$ *coincides with the group* $\mathrm{Aut}\,V$.

[1] As A. Tikhomirov has informed the author, paper [44] contains a mistake, but he knows a correct proof of the non-rationality of B_2.

For varieties of type A_2 or A_4 this was first proved in the joint paper by Yu. I. Manin and the author of this report [17] (see also [21]). For varieties of type B_1, the proof was begun by the author in [21] and finished by his student S. Hashin [16].

For the other Fano varieties $\operatorname{Bir} V \neq \operatorname{Aut} V$ and the simplest birational automorphisms not contained in $\operatorname{Aut} V$ are linked with curves of small degrees or with points on V (in the sense that these curves and points are the singularities of maximal multiplicities for the linear systems that define the birational automorphisms: such curves and points are called the *maximal singularities* of the corresponding birational automorphisms). By now the group $\operatorname{Bir} V$ has been described for varieties of types A_4', A_6 and B_2 (except for A_2, A_4 and B_1 above).

Now we shall formulate the results obtained.

PROPOSITION 1. (i) *Let V be a Fano variety of the type A_4'. Then to every line $Z \subset V$ (i.e., to every curve Z with $(Z - K_V) = 1$) not contained in the ramification divisor of the double covering $V \to Q = \boldsymbol{P}^4$ one can attach the birational involution a_Z defined by the linear systems*

$$a_Z(H) = 9H - 10Z - 6Z'$$

where Z' is the line conjugate to Z with respect to the double covering $V \to Q$.

(ii) *Let V be a Fano variety of type A_6. Then to every line $Z \subset V$ one can attach the birational involution a_Z defined by the linear system*

$$a_Z(H) = 4H - 5Z,$$

and to every conic $Y \subset V$ whose plane $\boldsymbol{P}(Y)$ lies in the quadric passing through $V \subset \boldsymbol{P}^5$ one can attach the birational involution β_Y defined by the linear system

$$\beta_Y(H) = 13H - 14Y - 8Z,$$

where Z is the line such that $Y \cup Z = V \cap P(Y)$.

(iii) *Let V be a Fano variety of type B_2. Then it has the following simplest birational involutions:*

a_P is the birational involution attached to a point $P \in V$ not on the ramification divisor of the double covering $\pi\colon V \to \boldsymbol{P}^3$ and defined by the linear system

$$a_P(H) = 3H - 4P;$$

β_C is the birational involution attached to a rational curve C such that $\pi(C)$ is the twisted cubic curve in $\boldsymbol{P}^3$ and defined by the linear system

$$\beta_C(H) = 9H - 5C - 3C',$$

where C' is the curve conjugate to C with respect to the double covering π: $V \to \boldsymbol{P}^3$;

γ_{C,C_1} *is the birational involution attached to a line $C \subset V$ (i.e., to a curve C with $(C \cdot H) = 1$) and to a rational curve C_1 which is a section of the pencil of hypersurfaces $|H - C - C'|$ and defined by the linear system*

$$\gamma_{C,C_1}(H) = (m+4)H - (m+2)C - mC' - 3C_1 - C_1',$$

where $m = (C_1 \cdot W) + \#(C \cap C_1) - 1$ and the mark $'$ stands, as before, for conjugation;

δ_{C,C_1} *is the birational involution attached to a line $C \subset V$ and a curve C_1 having a twofold intersection with the pencil of surfaces $|H - C - C'|$ and such that $C_1 = C_1'$ and defined by the linear system*

$$\delta_{C,C_1}(H) = (m+9)H - (m+4)C - mC' - 6C_1,$$

where $m = 3(C_1 \cdot H) - 6$;

ε_{C,C_1} *is the birational involution attached to a line C and a curve C_1 which also has a twofold intersection with the pencil $|H - C - C'|$ but $C_1 \neq C_1'$ and defined by the linear system*

$$\varepsilon_{C,C_1}(H) = (m+10)H - (m+5)C - mC' - 6C_1,$$

where m is some integer;

ξ_{C,C_1} *is the birational involution attached to a curve C such that $\pi(C)$ is a line in $\boldsymbol{P}^3$ and $C = \pi^{-1}\pi(C)$ and a curve C_1 which is a section of the pencil of hypersurfaces $|H - C|$ and defined by the linear system*

$$\xi_{C,C_1}(H) = (m+3)H - mC - 4C_1,$$

where $m = 6(C_1 \cdot H) + 6$.

THEOREM 6. *Let V be an arbitrary smooth variety of type A_4', A_6 or B_2. Let $B(V)$ denote the group of birational automorphisms generated by the involutions listed in Proposition* 1, (i), (ii) *or* (iii). *Then* Bir V *can be represented as a semidirect product of* Aut V *and the normal subgroup $B(V)$, i.e.* Bir V *can be represented as the extension*

$$1 \to B(V) \to \mathrm{Bir}\, V \to \mathrm{Aut}\, V \to 1$$

COROLLARY. *Any Fano threefold of one of the types A_4', A_6 and B_2 is non-rational.*

For varieties of types A_4' and A_6 these results have been obtained by the author [21] and for B_2 by S. Hashin (to appear in *Izv. Acad. Nauk SSSR*).

Remark 2. For a smooth three-dimensional cubic A_3, S. Tregub has come to similar but incomplete results. The birational automorphisms of Fano varieties of types A_8, A_{10} and, of course, of $\boldsymbol{P}^3$ have not been described yet. It would be interesting to apply this technique in studying the group $\mathrm{Aut}\boldsymbol{A}^3$, where $\boldsymbol{A}^3$ is the affine space of dimension three.

4. Now we pass to the consideration of the second class of threefolds, i.e., conic bundles. Here we assume char $k \neq 2$ unless otherwise stated.

DEFINITION 2. Let V be a smooth projective irreducible threefold and S a smooth projective surface. A surjective morphism π: $V \to S$ is called a *standard conic bundle* if it satisfies the following conditions:

(i) for each point $s \in S$ the scheme-theoretic fibre $V_s = \pi^{-1}(S)$ is isomorphic (as a scheme) over $k(s)$ to a conic (possibly degenerate) in $\boldsymbol{P}^2_{k(s)}$, where $k(s)$ is the residue field of the point;

(ii) (relative minimality) for each irreducible curve $D \subset S$ the surface $V_D = \pi^{-1}(D)$ is irreducible over k.

We are going to mention some known results concerning conic bundles.

PROPOSITION 2. (i) *Let* π: $V \to S$ *be a standard conic bundle, and* $C \subset S$ *its discriminant locus* (*possibly* $C = \emptyset$). *Then* C *is a reduced divisor with normal crossings and*

$$V_S \cong \begin{cases} \boldsymbol{P}^1, & \text{if } s \in S\setminus C; \\ \boldsymbol{P}^1 \vee \boldsymbol{P}^1, & \text{two intersecting lines if } s \in C - \mathrm{Sing}\, C; \\ 2\boldsymbol{P}^1, & \text{a double line if } s \in \mathrm{Sing}\, C. \end{cases}$$

Here Sing C *is the set of double points of* C (see [3], [38], [52]).

(ii) *Under condition* (i) *there is an isomorphism*

$$\mathrm{Pic}\, V \simeq \pi^* \mathrm{Pic}\, S + \begin{cases} \boldsymbol{Z}[K_v] & \text{if } \pi\colon V \to S \text{ has no rational sections}, \\ \boldsymbol{Z}[L] & \text{if } \pi\colon V \to S \text{ has rational sections}, \end{cases}$$

where L *is the class of a rational section* $S \to V$.

(iii) *Let* S *be a rational surface,* $l \neq 2$ *a prime, and* $H^i(V, \boldsymbol{Z}_l)$ *the* l*-adic cohomology. Then*

$$H^0(V, \boldsymbol{Z}_l) \simeq H^6(V, \boldsymbol{Z}_{l},) \simeq \boldsymbol{Z}_l, \quad H^1(V, \boldsymbol{Z}_l) \simeq H^5(V, \boldsymbol{Z}_l) = 0,$$

$$H^3(V, \boldsymbol{Z}_l) \simeq \bigoplus_{i=1}^{r} H^1(\tilde{C}_i, \boldsymbol{Z}_l)/H^3(C_i, \boldsymbol{Z}_l) \oplus (\boldsymbol{Z}/2\boldsymbol{Z})^{c-1},$$

where c *is the number of connected components of the curve* C, C_i *is an irreducible component of* C $(1 \leqslant i \leqslant r)$, $\tilde{C}_i$ *is the curve parametrizing the components of the fibres of the surface* $V_{C_i} = \pi^{-1}(C_i)$ (see [52]).

(iv) *Let* γ: $W \to T$ *be a rational map*, $\dim W = 3$, $\dim T = 2$ *such that the generic fibre* W_η *is a curve of genus zero over* $k(\eta)$. *Then there exists a standard conic bundle* π: $V \to S$ *birationally equivalent to* γ: $W \to T$ (see [52], [38]).

Since we are interested mainly in rationality questions, we shall assume from now on that S is a rational surface.

THEOREM 7 (see [1]). *Under the conditions of Proposition* 2, (iii), *we have: if* $c > 1$, *then* V *is non-rational.*

Remark 3. This theorem is proved by using the birational invariant $\operatorname{Tors} H^3(V, \boldsymbol{Z}) = \operatorname{Br} V$, where $\operatorname{Br} V$ denotes the Brauer–Grothendieck group of V (see [1]).

If $c = 1$ then, as in the case of Fano threefolds, to prove the non-rationality of the conic bundle V two methods are used, namely: the Clemens–Griffiths intermediate Jacobian method and the untwisting birational automorphism method (the first of these two methods works over $\boldsymbol{C}$).

PROPOSITION 3. *Let* $k = \boldsymbol{C}$ *and let* π: $V \to S$ *be a standard conic bundle,* S *a rational surface, and* $C \subset S$ *the discriminant curve with* $c = 1$. *Let* $\tilde{\pi}$: $\tilde{C} \to C$ *be the double covering associated with* π: $V \to S$, *where* $\tilde{C}$ *is the curve of components of fibres of the surface* $V_C = \pi^{-1}(C)$, $\operatorname{Pr}(\tilde{C}/C)$ *the Prym variety* (see [2], [42]) *associated with the covering* $\tilde{\pi}$: $\tilde{C} \to C$ *and* $J(V)$ *the intermediate Jacobian of* V. *Then* $\operatorname{Pr}(\tilde{C}/C)$ *is a principally polarized abelian variety and there is an isomorphism* $\operatorname{Pr}(\tilde{C}/C) \simeq J(V)$.

In the case $S = \boldsymbol{P}^2$ this has been proved by Beauville [3] and in the general case by S. Endryushka [8] (see also Beltrametti and Francia [53]).

In order to prove the non-rationality of V one can now use the following result about distinguishing Prymians from Jacobians.

THEOREM 8 (Mumford [27], Beauville [2], Shokurov [42]). *Let* $\operatorname{char} k \neq 2$, *and let* $(\tilde{C}, I)$ *be a pair consisting of a connected curve* $\tilde{C}$ *with at most ordinary double points as singularities and an involution* I *on* $\tilde{C}$ *satisfying the Beauville conditions* [2]:

$$\text{(B)}\quad I(\operatorname{Sing}\tilde{C}) = \operatorname{Sing} C, \qquad \operatorname{Fix} I \stackrel{\text{def}}{=} \{x \in \tilde{C} \mid I(x) = x\} = \operatorname{Sing}\tilde{C},$$

where $C = \tilde{C}/I$ *is the quotient curve, and* Sing *denotes the set of singular points. Suppose that one more condition is satisfied* [42]:

(S) *for any decomposition* $\tilde{C} = \tilde{C}_1 \cap \tilde{C}_2$ *we have* $\#\{\tilde{C}_1 \cap \tilde{C}_2\} \geqslant 4$.

Then $\mathrm{Pr}(\tilde{C}/C)$ *is isomorphic to the Jacobian (or to the product of Jacobians) of a nonsingular curve as a principally polarized abelian variety if and only if the curve* $C = \tilde{C}/I$ *is one of the following:*

(a) C *is a hyperelliptic curve, i.e., there exists a finite morphism* $C \to \boldsymbol{P}^1$ *of degree* 2;

(b) C *is a trigonal curve, i.e., there exists a finite morphism* $C \to \boldsymbol{P}^1$ *of degree* 3;

(c) C *is a quasitrigonal curve, i.e., is a hyperelliptic curve with two points* P *and* Q, $P+Q \notin G_1^2$, *identified;*

(d) $C \subset \boldsymbol{P}^2$ *is a quintic curve and* $h^0\big(\tilde{C}, \tilde{\pi}^* O_c(1)\big) = 3$.

COROLLARY. *Under the assumption of Proposition* 3, *if* $\tilde{C}$ *satisfies condition* (S) (*condition* (B) *is satisfied automatically*) *and if the curve* C *belongs to none of the types* (a), (b), (c), (d) *of Theorem* 8, *then the variety* V *is non-rational. In particular, if* $S = \boldsymbol{P}^2$ *and* $\deg C \geqslant 6$ *then* V *is non-rational* (Beauville [3]).

Remark 4. We have seen that the intermediate Jacobian method enables us to prove the non-rationality of a large class of those threefolds which admit the structure of a conic bundle. Just in this way S. Endryushka has proved the non-rationality of a general Enriques variety [9] (see also [31]): this variety can be birationally represented as a conic bundle with a curve $C \subset S$ of genus 5 which is not a curve of type (a), (b), (c) or (d) from Theorem 8. Do there exist non-rational three-dimensional conic bundles whose intermediate Jacobian is the Jacobian of a curve? The affirmative answer to this question is given below.

5. Let us now consider the method of the untwisting birational automorphisms applied to standard conic bundles.

THEOREM 9 (V. G. Sarkisov [37], [38]). *Let* $\mathrm{char}\, k \neq 2$ *and let* $\pi\colon V \to S$ *be a standard conic bundle where* S *is a rational surface and* $\chi\colon V \to V$ *an arbitrary birational map. Then, if* $4K_S + C \geqslant 0$ (K_S *is the canonical divisor on* S), *there exists such a birational isomorphism* $\alpha\colon S \to S$ *that* $\alpha \circ \pi = \pi \circ \chi$, *i.e.,* χ *maps fibres of* π *into fibres.*

COROLLARY 1. *Under the assumptions of Theorem* 9 *the group* $\mathrm{Bir}\, V$ *can be represented in the form of an extension*

$$1 \to \mathrm{Aut}\, V_\eta \to \mathrm{Bir}\, V \to G,$$

where V_η *is the generic fibre of the morphism* π *over the residue field* $k(\eta)$ *of the generic point* $\eta \in S$ *and* $G \subset \mathrm{Bir}\, S$ *is some subgroup.*

COROLLARY 2. *Under the assumptions of Theorem* 9 *the variety* V *is non-rational.*

The question of the existence of standard conic bundles with given degenerations can be solved by using the bijective correspondence between generic fibres V_η of conic bundles and quaternion algebras A_η over $k(\eta)$ (see [1]). Via this correspondence any standard conic bundle becomes, over S, a Severi–Brauer variety associated with some maximal order in the corresponding algebra A_η. In that way, due to the exact sequence of Artin–Mumford [1],

$$0 \to \operatorname{Br} S \to \operatorname{Br} k(S) \to \bigoplus_{\text{curves } C_i \subset S} H^1(k(C_i), Q/Z) \to \bigoplus_{\text{points } x \in \nabla C_i} \mu_x^{-1} \to \mu^{-1} \to 0$$

where $\mu^{-1} = \bigcup_n \mu_n^{-1} = \bigcup_n \operatorname{Hom}(\mu_n, Q/Z)$, μ_n is the group of roots of the nth degree of 1, the question is reduced to the question of the existence of quaternion algebras with given invariants (up to $\operatorname{Br} S$). If S is a rational surface, then $\operatorname{Br} S = 0$ and all elements from $\operatorname{Br} k(S)$, particularly elements of the second order, are uniquely determined by their invariants. It turns out (see [38]) that in the case where $\dim S = 2$ all elements of the second order are represented by quaternion algebras. This brings us to the following result.

Proposition 4. *Let S be a rational surface over k,* $\operatorname{char} k \neq 2$, *$C \subset S$ a curve with normal crossing and $\tilde{\pi}$: $\tilde{C} \to C$ a double covering which satisfies the Beauville condition* (B). *Then there exists a standard conic bundle π: $V \to S$ having the discriminant curve C and such that the components of fibres of the surface $V_C = \pi^{-1}(C)$ are parametrized by $\tilde{C}$.*

From Theorem 9 and Proposition 4 one can easily derive the following

Corollary. *There exist non-rational standard conic bundles π: $V \to S$ over a rational surface S such that $H^3(V, Z_l) = 0$.*

To construct such a conic bundle one can take for example an irreducible curve with ordinary singularities $C' \subset P^2$ with the birational genus 1 and $\deg C' \geqslant 12$. Let $S \to P^2$ be the blow-up of the singular points of C', and C the proper transform of C'. Then C is a smooth elliptic curve. Let $\tilde{\pi}$: $\tilde{C} \to C$ be one of the non-trivial unramified double coverings; then the conic bundle which exists by Proposition 4 is non-rational by Theorem 9 and, as one can easily verify, $H^3(V, Z_l) = 0$. In particular, there exist non-rational three-dimensional varieties with trivial intermediate Jacobians over C.

6. The following two criteria of the rationality of standard conic bundles have been conjectured by the present author [22] and V. V. Shokurov [42].

Conjecture I (see [22]). *Let* $\operatorname{char} k \neq 2$ *and let π: $V \to S$ be a standard* *conic bundle over a rational surface S with a connected degeneracy curve $C \subset S$ (possibly $C = \emptyset$). Then the following two conditions are equivalent:*

(i) *V is rational;*

(ii) *one of the following two statements is true:*

(a) *There exists a pencil of rational curves $\{L_\lambda, \lambda \in \boldsymbol{P}^1\}$ without fixed components on S such that $(L_\lambda \cdot C) \leqslant 3$;*

(b) *there exists a birational map φ: $S \to \boldsymbol{P}^2$ such that $C^1 \approx \varphi(C)$ is a curve of degree 5 with normal crossing and for the double cover $\tilde{\pi}$: $\tilde{C}' \to C'$ induced by π: $V \to S$ the condition $h^0(\tilde{C}', \tilde{\pi}^* O_{C'}(1)) = 3$ is fulfilled.*

The implication (ii)$\to$(i) is proved in [22]. The converse is proved in the same paper under an extra condition: *if V is rational, then there exists a birational map β: $V \to \boldsymbol{P}^3$ which maps fibres of the morphism π onto conics in $\boldsymbol{P}^3$.*

Thus Conjecture I is reduced to the following.

CONJECTURE I′ (see [22]). *Every two-dimensional system (of index 1) of rational curves on $\boldsymbol{P}^3$ can be reduced to a two-dimensional system of conics by means of birational transformations.*

CONJECTURE II (see [42]). *Under the assumption of Conjecture* I *the following conditions are equivalent:*

(i) *V is rational;*

(ii) *one of the following two statements is true:*

(a) *the linear system $|2K_S + C|$ is empty;*

(b) *the same as condition* (b) *in Conjecture* I.

This conjecture has been proved in the case where $S \simeq \boldsymbol{P}^2$ or $S = \boldsymbol{F}_N$, i.e., S is a geometric ruled surface (see [42]).

7. To conclude, we mention some of the open problems.

(i) It is not known if every standard conic bundle V with a rational base S is unirational. Only some special examples of unirational conic bundles are known (see [1], [30], [37]). Presumably a general variety of this kind is not unirational. Note that the unirationality of $V\,(\pi$: $V \to S)$ is equivalent to the existence of a rational surface $T \subset V$ such that $\pi|_T$: $T \to S$ is finite at the generic point (see e.g. [36]).

(ii) In the birational classification programme of three-dimensional varieties with Kodaira dimension $-\infty$ it is assumed that every such variety has a minimal model which admits terminal singularities and which belongs to one of the following types (modulo a finite list of exceptions each of which can be treated separately) see [33], [34], [24]:

(a) Fano varieties with at most terminal singularities;
(b) conic bundles with at most terminal singularities;
(c) pencils of del Pezzo surfaces, also with at most terminal singularities.
The problem of a characterization of rational varieties from each of these classes is actual.

(iii) Little is known about the rationality and non-rationality of varieties of dimension greater than four. Theorem 9 has been extended by V. G. Sarkisov [38] to conic bundles over a rational base of an arbitrary dimension. As before, the question of the rationality of a general cubic hypersurface of dimension at least four remains open. Only special examples of rational smooth cubic hypersurfaces of an even dimension $\geqslant 4$ are known (see [36], [49]).

References

[1] Artin M. and Mumford D., Some elementary examples of unirational varieties which are non-rational, *Proc. London Math. Soc.* **25** (1972), pp. 75–95.

[2] Beauville A., Prym varieties and the Schottky problem, *Invent. Math.* **41** (1977), pp. 149–196.

[3] Beauville A., Varieties de Prym et Jacobiennes intermediaires, *Ann. Sci. Ecole norm. Super.* **10** (1977), pp. 309–391.

[4] Beauville A., Les singularities du diviseur Θ de la Jacobienne intermediaire de l'hypersurfase cubique dans P^4, *Lect. Notes Math.* **947** (1982), pp. 190–208.

[5] Bloch S. and Murre J. P., On the Chow group of certain types of Fano threefolds, *Comp. Math.* **39** (1979), pp. 47–105.

[6] Clemens C. H., Degeneration techniques in the study of threefolds, *Lect. Notes Math.* **947** (1982), pp. 93–154.

[7] Clemens C. H. and Griffiths P. A., The intermediate Jacobians of the cubic threefolds, *Ann. of Math.* **95** (1972), pp. 281–356.

[8] Эндрюшка С., Средний якобиан трёхмерного расслоения на коники является многообразием Прима, *Матем. сб.* **122** (2) (1983), 276–285.

[9] Эндрюшка С., Нерациональность общего многообразия Энриквеса, *Матем. сб.* 1984 (in press).

[10] Fano G., Osservazioni sopra alcune varietà non razionali aventi tutti i generi nulli, *Atti Acad. Torino* **50** (1915), pp. 1067–1072.

[11] Fano G., Sulle varietà algebriche a tre dimensioni aventi tutti i generi nulli, *Atti Cong. Internaz. Bologna* **4** (1931), pp. 115–119.

[12] Fano G., Sulle varietà a tre dimensioni curve — sezioni canociche, *Mem. R. Acad. d'Italia* **8**, (1937), pp. 23–64.

[13] Fano G., Nuove ricerche sulle varietà algebriche a tre dimensioni a curve sezioni canociche, *Comm. Pont. Acc. Sc.* **11** (1947), pp. 635–720.

[14] Гушель Н. П., О многообразиях Фано рода 6, *Изв. АН СССР, Сер. Матем.* **46.6** (1982), pp. 1159–1174.

[15] Гушель Н. П., О многообразиях Фано рода 8, *УМН* **38.1** (1983), pp. 163–164.

[16] Хашин С. И., Бирациональные автоморфизмы двойного конуса Веронезе размерности 3, *Вестник МГУ. Сер. Матем., механика*, N1 (1984), pp. 13–16.

[17] Исковских В. А., Манин Ю. И., Трёхмерные квартики и контрпримеры к проблеме Люрота, *Матем. сб.* **86.1** (1971), pp. 140–166.

[18] Исковских В. А., Трёхмерные многообразия Фано. I, *Изв. АН СССР. Сер. матем.* **41.3** (1977), pp. 516–562.

[19] Исковских В. А., Трёхмерные многообразия Фано. II, *Изв. АН СССР. Сер. матем.* **42.3** (1978), pp. 504–549.

[20] Исковских В. А., Антиканонические модели трёхмерных алгебраических многообразий, *Совр. проблемы матем., ВИНИТИ* **12** (1979), pp. 59–157.

[21] Исковских В. А., Бирациональные автоморфизмы трёхмерных алгебраических многообразий, *Совр. проблемы матем., ВИНИТИ* **12** (1979), pp.159–236.

[22] Исковских В. А., Конгруэнции коник в P^3, *Вестник МГУ. Сер. матем., механика* **6** (1982), pp. 57–62.

[23] Iskovskikh V. A. and Shokurov V. V., Biregular theory of Fano 3-folds, *Proc. Alg. Geom. Conf., Copenhagen*, **1978**; *Lect. Notes Math.* **732** (1979), pp. 171–182.

[24] Kawamata Y., *Elementary construction of algebraic 3-folds*, preprint, 1983.

[25] Mori S., Threefolds whose canonical bundles are not numerically effective, *Ann. of Math.*, **116** (1982), pp. 133–176.

[26] Mori S. and Mukai S., Classification of Fano 3-folds with $B_2 \geqslant 2$, *Manuscr. Math.* **36.2** (1981), pp. 147–162.

[27] Mumford D., *Prym varieties I, Contrib. to Analysis*, Academic Press, 1974, pp. 325–350.

[28] Murre J. P., Reduction of the proof of the nonrationality of nonsingular cubic to a result of Mumford, *Comp. Math.* **27.1** (1973), pp. 63–82.

[29] Murre J. P., *Classification of Fano threefolds according to Fano and Iskovskih*, Lect. Notes Math. **947** (1982), pp. 35–92

[30] Botta P. and Verra A., An example of unirational nonrational threefold with a conic bundle over a rational surface, *Proc. Kon. ned. acad. Wetensch.* **83.1** (1980), pp. 19–32.

[31] Botta P. and Verra A., The non-rationality of generic Enriques threefold, *Comp. Math.* **48** (2) 1983, pp, 167–184.

[32] Puts P. J., On some Fano-threefolds that are sectioned of Grassmannians, *Proc. Konikl. Nederl. Acad. Wet., Ser. A*, **85.1** (1982), pp. 77–80.

[33] Reid M., Canonical 3-folds, in: *Algebraic geometry, Angers* **1979**, *Netherlands*, 1980, pp. 273–310.

[34] Reid M., *Minimal models of canonical 3-folds*, Preprint Kyoto Univ., 1981, pp. 1–55.

[35] Reid M., *Lines on Fano 3-folds according to Shokurov*, Mittag–Leffler Report, No **11**, 1980.

[36] Roth L., *Algebraic threefolds with special regard to problem of rationality*, Springer Verlag, Berlin-Göttingen-Heidelberg, 1955.

[37] Саркисов В. Г., Бирациональные автоморфизмы расслоений на коники, *Изв. АН СССР, Сер. матем.* **44.4** (1980), pp. 918–944.

[38] Саркисов В. Г., О структурах расслоений на коники, *Изв. АН СССР. Сер. матем.* **46.2** (1982), pp. 371–408.

[39] Segre B., Variazione continus ed omotopis in geometrie algebrice, *Ann. Mat. pura ed appl. Ser. IV*, L., 1960, pp. 149–186.
[40] Шокуров В. В., Гладкость общего антиканонического дивизора на многообразии Фано, *Изв. АН СССР. Сер. матем.* **43.2** (1979), pp. 430–441.
[41] Шокуров В. В., Существование прямой на многообразиях Фано, *Изв. АН СССР. Сер. матем.* **43.4** (1979), pp. 922–964.
[42] Шокуров В. В., Многообразия Прима: теория и приложения, *Изв. АН СССР. Сер. матем.* **47.4** (1983), pp. 785–856.
[43] Тихомиров А. С., Геометрия поверхности Фано двойного пространства P^3 с ветвлением в квартике, *Изв. АН СССР. Сер. матем.* **44.2** (1980), pp. 415–442.
[44] Тихомиров А. С., Средний якобиан двойного пространства P^3, разветвлённого в квартике, *Изв. АН СССР. Сер. матем.* **44.6** (1980), pp. 1329–1377.
[45] Тихомиров А. С., Поверхность Фано двойного конуса Веронезе, *Изв. АН СССР. Сер. матем.* **45.5** (1981), pp. 1121–1197.
[46] Тюрин А. Н., Пять лекций о трёхмерных многообразиях, *УМН* **27.5** (1972), pp. 3–50.
[47] Тюрин А. Н., Средний якобиан трёхмерных многообразий, *Совр. проблемы матем., ВИНИТИ* **12** (1979), pp. 5–57.
[48] Тюрин А. Н., О пересечении квадрик, *УМН* **30.6** (1975), pp. 51–99.
[49] Трегуб С. Л., О некоторых конструкциях рациональности четырёхмерных кубик, *Вестник МГУ. Серия матем., механика*, 1984 (in press).
[50] Unumura H., *Maximal algebraic subgroups of the Cremona group of three variables — imprimitive algebraic subgroups of exceptional type*, 1981.
[51] Welters C. E., *Abel–Jacobi isogenics for certain types of Fano threefolds*, Amsterdam 1981.
[52] Загорский А. А., О трёхмерных конических расслоениях, *Мат. заметки* **21.6** (1977), pp. 745–758.

Added in proof:

[53] Beltrametti M. and Francia R., *Conic bundles on non-rational surfaces*, Preprint, 1983.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

SHIGEFUMI MORI*

Cone of Curves, and Fano 3-Folds

§ 1. Cones of curves, extremal rays, and Fano 3-folds

Through the solution of Frankel and Hartshorne conjectures (Mori [10], Siu and Yau [20]), the importance of finding a rational curve on a given algebraic variety is exibited. It is done through a cone of curves and extremal rays.

Let X be an n-dimensional non-singular projective variety defined over an algebraically closed field k of characteristic $p \geqslant 0$. Let $N_{\mathbf{Z}}(X)$ be the set of numerical equivalence classes of 1-cycles on X and let $N(X) = N_{\mathbf{Z}}(X) \otimes_{\mathbf{Z}} \mathbf{R}$. Then $N(X)$ is a real vector space of dimension $\varrho(X)$, and in $N(X)$ we consider the smallest closed convex cone $\overline{NE}(X)$ (the *cone of curves*) containing all the classes of irreducible curves which are stable under multiplication by $\mathbf{R}_+$, the set of non-negative real numbers. This cone was originally considered by Hironaka [3] and Kleiman [5]. A new notion here is the extremal ray, which is defined as follows. Let

$$\overline{NE}_-(X) = \{z \in \overline{NE}(X) \mid (z. -K_X) \leqslant 0\}.$$

We say that a half line $R = \mathbf{R}_+ z \subset \overline{NE}(X)$ is an *extremal ray* if there exists a closed convex cone $B \supset \overline{NE}_-(X)$ such that $R \not\subset B$ and $\overline{NE}(X) = R + B$.

THEOREM (Mori [11]). *$\overline{NE}(X)$ is the smallest closed convex cone containing $\overline{NE}_-(X)$ and all the extremal rays. An arbitrary extremal ray is generated by the class of a (possibly singular) rational curve C such that* $1 \leqslant (C. -K_X) \leqslant n+1$.

A divisor D on X is called *numerically effective* if $(D.C) \geqslant 0$ for an arbitrary irreducible curve C. An obvious corollary is

* Partially supported by the Japan–U.S.A. exchange foundation.

COROLLARY. *X has an extremal ray if and only if K_X is not numerically effective.*

As a special case of the theorem, one has

COROLLARY. *If $-K_X$ is ample, then $\overline{NE}(X)$ is spanned by a finite number of extremal rays.*

If $\dim X = 2$, or $\dim X = 3$ and $p = 0$, extremal rays are geometrically explained by means of "contractions".

THEOREM—DEFINITION. *Assume that* $\dim X = 2$, *or* $\dim X = 3$ *and* $p = 0$. *For an arbitrary extremal ray R of X, there exists a morphism $f\colon X \to Y$ to a projective variety such that*

(i) $f_* \mathcal{O}_X = \mathcal{O}_Y$ *and*

(ii) *an arbitrary irreducible curve C is collapsed to a point if and only if the class $[C]$ of C belongs to R. Such an f is unique up to an isomorphism and*

(iii) *$-K_X$ is f-ample.*

We call f the contraction of R.

For $\dim X = 2$, the description of the extremal ray is equivalent to the following classical result.

THEOREM. *Assume that* $\dim X = 2$, *and X has an extremal ray R. Let $f\colon X \to Y$ be the contraction of R. Then one of the following hold.*

(i) *f is the contraction of an exceptional curve C of the first kind, and $R = \boldsymbol{R}_+[C]$.*

(ii) *f is a $\boldsymbol{P}^1$-bundle, and R is generated by a fiber of f.*

(iii) *$X \simeq \boldsymbol{P}^2$, $Y \simeq \operatorname{Spec} k$, and $R = \overline{NE}(X)$.*

The result for $\dim X = 3$ and $p = 0$ turns out to be a natural generalization of the above.

THEOREM. *Assume that* $\dim X = 3$, $p = 0$, *and X has an extremal ray R. Let $f\colon X \to Y$ be the contraction of R. Then $\varrho(X) = \varrho(Y)+1$, and one of the following holds.*

(i) *f contracts an irreducible divisor D to a curve or to a point, f is then the blow-up of Y along the reduced closed subscheme $f(D)$, and one of the following 5 cases holds:*

(i.1) *Y is non-singular, $f(D)$ is a non-singular curve, D is a $\boldsymbol{P}^1$-bundle over $f(D)$, and $\mathcal{O}_D(-D)$ is the tautological line bundle,*

(i.2) *Y is non-singular, $f(D)$ is a point, $D \simeq \boldsymbol{P}^2$, and $\mathcal{O}_D(-D) \simeq \mathcal{O}_P(1)$,* (i.3) *$f(D)$ is an ordinary double point of Y such that $\mathcal{O}_{Y,f(D)}$ is factorial, $D \simeq \mathbf{P}^1 \times \mathbf{P}^1$, and $\mathcal{O}_D(-D) \simeq \mathcal{O}_D(1,1)$,*

(i.4) *$f(D)$ is a double point of Y, D is isomorphic to an irreducible singular quadric surface Q of $\mathbf{P}^3$, and $\mathcal{O}_D(-D) \simeq \mathcal{O}_Q(1)$, and*

(i.5) *$f(D)$ is a quadruple point, $D \simeq \mathbf{P}^2$, and $\mathcal{O}_D(-D) \simeq \mathcal{O}_{\mathbf{P}}(2)$.*

(ii) *f is flat and one of the following holds:*

(ii.1) *an arbitrary fiber of f is isomorphic to a conic in $\mathbf{P}^2$ as a scheme (f is called a conic bundle), and*

(ii.2) *an arbitrary fiber of f is an irreducible reduced surface D such that ω_D^{-1} is ample (f is called a del Pezzo fibring).*

(iii) *$Y \simeq \operatorname{Spec} k$, hence $\varrho(X) = 1$ and $-K_X$ is ample (cf.* (1), (2) *and Iskovskih's classification in* (3), (4) *of the next theorem).*

We say that an extremal ray $R = \mathbf{R}_+ z$ is *numerically effective* if $(z.D) \geqslant 0$ for every irreducible divisor D on X. Then R is numerically effective exactly in cases (ii), (iii) of the above theorem.

We say that X is a *Fano n-fold* if $-K_X$ is ample. The index r of a Fano 3-fold X is, by definition, the greatest integer (>0) such that $-K_X \in r\operatorname{Pic} X$.

The following results are known in characteristic 0:

THEOREM.

(0) *One has $r = 1, 2, 3,$ or 4,*

(1) *if $r = 4$, then $X \simeq \mathbf{P}^3$* ([8]),

(2) *if $r = 3$, then $X \simeq$ 3-dimensional quadric* ([8]),

(3) *if $r = 2$, and $\varrho(X) = 1$ then there are exactly* 5 *deformation types of X (Iskovskih* [4] *and Fujita* [1], *cf.* [2] *for* char p),

(4) *if $r = 1$ and $\varrho(X) = 1$ then Iskovskih* [4] *classified X's into* 10 *families using the results of Shokurov* [18], [19] *(we refer the reader to Iskovskih's report in this volume for the precise classification of* (3) *and* (4)) *and*

(5) *If $r = 1$ and $\varrho(X) \geqslant 2$ then there are exactly* 87 *deformation types, Mori–Mukai* [12]:

$\varrho(X)$	2	3	4	5	6	7	8	9	10	$\geqslant 11$
# of def. types	36	31	12	3	1	1	1	1	1	0

By a technique similar to the one in [10], Janos Kollar proved

THEOREM. *In any characteristic an arbitrary Fano n-fold X is uni-ruled. In other words, given any point x of X, there exists a rational curve C through x such that $(C. -K_X) \leqslant n+1$.*

§ 2. Terminal singularities

To continue the process of contraction of extremal rays, one has to consider projective varieties with "terminal singularities" in the sense of Reid [15].

DEFINITION (Reid [15]). Let $k = \boldsymbol{C}$. Let y be a point of a normal 3-fold Y. Then we say that Y has only *terminal singularity* (resp. *canonical singularity*) at y if there is a natural number r such that the Weil divisor $r K_Y$ is actually a Cartier divisor and, for some resolution (or equivalently, for an arbitrary resolution) $f\colon Y' \to Y$, the sections of $\mathcal{O}(r K_Y)$ vanish along every divisor in $f^{-1}(y)$ (resp. are regular in a neighborhood of $f^{-1}(y)$) when considered as meromorphic sections of $r K_{Y'}$.

THEOREM (Reid [15]). *If (Y, y) is a 3-dimensional terminal singularity, then (Y, y) is the quotient, by a cyclic group action, of an isolated cDV singularity. Such a singularity is, by definition, an isolated 3-dimensional hypersurface singularity (Z, z) whose general hyperplane section through z has a rational double (or smooth) point at z.*

DEFINITION. Assume that $k = \boldsymbol{C}$. We say that a normal algebraic variety X is *$\boldsymbol{Q}$-factorial at x* if the divisor class group $\mathrm{Cl}(\mathcal{O}_{X,x})$ of the local ring $\mathcal{O}_{X,x}$ at x is torsion, and that X is *$\boldsymbol{Q}$-factorial* if X is $\boldsymbol{Q}$-factorial at every point (Reid [16]). Let $f\colon X \to Y$ be a projective morphism such that X is a $\boldsymbol{Q}$-factorial 3-fold with only terminal singularities and $f_* \mathcal{O}_X = \mathcal{O}_Y$. Let us consider 1-cycle $\sum a_C C$ which is a linear combination of curves C such that $f(C)$ is a point. Let $N_{\boldsymbol{Z}}(X/Y)$ be the set of numerical equivalence classes of such 1-cycles, $N(X/Y) = N_{\boldsymbol{Z}}(X/Y) \otimes_{\boldsymbol{Z}} \boldsymbol{R}$ and $\overline{NE}(X)$ be the smallest closed convex cone containing all the classes of irreducible curves C on X such that $f(C)$ is a point. We say that a half line $R \subset \overline{NE}(X/Y)$ is an *extremal ray of X/Y* (or an *f-extremal ray*) if R is generated by the class of a curve and there is a closed convex cone B of $\overline{NE}(X/Y)$ such that $B \not\supset R$, $\overline{NE}(X/Y) = R + B$, and $B \supset \overline{NE}_-(X/Y)$, where

$$\overline{NE}_-(X/Y) = \{z \in \overline{NE}(X/Y) \mid (z.\ -K_X) \leqslant 0\}.$$

We fix the above notation. One can ask the following

QUESTION. Is $\overline{NE}(X/Y)$ the smallest closed convex cone containing $\overline{NE}_-(X/Y)$ and all the f-extremal rays? For an arbitrary open convex cone U containing $\overline{NE}_-(X/Y)$, are there only a finite number of f-extremal rays which are not contained in $U \cup \{0\}$?

The answer is affirmative if $Y = \operatorname{Spec} k$ and K_X is pseudo-effective (this is the case if $\varkappa(X) \geqslant 0$) (Kawamata [7]), or if $\dim Y \geqslant 1$ (done essentially by Tsunoda [21]).

If X/Y has an extremal ray R, and if R is not numerically effective, one contracts the extremal ray (Kawamata [7]), but if the contracted 3-fold X' is not Q-factorial, we have to blow up X' to get a better model (and finally to get a "minimal" model). However, this process is not well understood yet. Possibly one should say that X/Y is a *minimal model* if X/Y does not have an extremal ray which is not numerically effective. If $\dim Y = 0$ and K_X is pseudo-effective, then this is equivalent to saying that K_X is numerically effective (Reid [15]). Kawamata [6] proved that if $\dim Y = 0$ and X is a minimal model (over Y) of general type then the canonical ring $\bigoplus_m H^0(X, \mathcal{O}(mK_X))$ is finitely generated.

Let $\dim Y = 1$ and assume that X/Y has everywhere semi-stable reduction with general fiber X_t. It is known that X/Y has a minimal model if X_t is a $K3$ surface or an abelian surface (Kulikov [9], Persson and Pinkham [14]) or if X_t is an Enriques' surface or a hyperelliptic surface with $2K \sim 0$ (Morrison [13] through combinatorial birational geometry). Tsunoda [21] seems to have proved that X/Y has a minimal model if X_t is a mimimal model with $\varkappa(X) \geqslant 0$ by using the logarithmic version of cones of curves and extremal rays. If a three-dimensional and Q-factorial Y has only terminal singularities, then the process of finding a minimal model over Y is the same as factoring the birational morphism $X \to Y$ into the product of elementary birational maps. However, this process is not well understood yet.

After having written up this paper, the author found that the question marked here was also put as a conjecture in the newly added part of [7].

References

[1] Fujita T., On the Structure of Polarized Manifolds With Total Deficiency One, I and II, *J. Math. Soc. Japan* **32** (1980), pp. 709–725, and **33** (1981), pp. 415–434.

[2] Fujita T., On Polarized Varieties of Small Δ-Genera, *Tôhoku Math. Journ.* **34** (1982), pp. 319–341.

[3] Hironaka H., *On the Theory of Birational Blowing-Up* (Thesis), Harvard, 1960.

[4] Iskovskih V. A., Fano 3-Folds, I and II, *Izv. Akad. Nauk SSSR Ser. Mat.* **41** (1977) and **42** (1978), English transl.: *Math. USSR Izv.* **11** (1977), pp. 485–527 and **12** (1978), pp. 469–506.

[5] Kleiman S., Toward a Numerical Theory of Ampleness, *Ann. of Math.* **84** (1966), pp. 293–344.

[6] Kawamata Y., *On the Finiteness of Generators of a Pluri-Canonical Ring for a 3-Fold of General Type*, to appear.

[7] Kawamata Y., *Elementary Contractions of Algebraic 3-Folds*, to appear.
[8] Kobayashi S. and Ochiai T., Characterization of Complex Projective Spaces and Hyperquadrics, *J. Math. Kyoto U.* **13** (1973), pp. 31–47.
[9] Kulikov V. S., Degenerations of K-3 Surfaces and Enriques' Surfaces, *Izv. Akad. Nauk SSSR Ser. Mat.* **41** (1977), English transl.: *Math. USSR Izv.* **11** (1977), pp. 957–989.
[10] Mori S., Projective Manifolds with Ample Tangent Bundles, *Ann. of Math.* **110** (1979), pp. 593–606.
[11] Mori S., Threefolds Whose Canonical Bundles are not Numerically Effective, *Ann. of Math.* **116** (1982), pp. 133–176.
[12] Mori S., and Mukai S., *Manuscripta Math.* **36** (1981), pp. 147–162.
[13] Morrison D., *Semistable Degenerations of Enriques' and Hyperelliptic Surfaces* (Thesis), Harvard, 1980.
[14] Persson U. and Pinkham H., Degeneration of Surfaces with Trivial Canonical Bundle, *Ann. of Math.* **113** (1981), pp. 45–66.
[15] Reid M., Minimal Models of Canonical 3-Folds, *Symp. Math.* **1**, Kinokuniya and North–Holland, 1982.
[16] Reid M., *Decomposition of Toric Morphisms*, to appear.
[18] Shokurov V. V., The Smoothness of the General Anticanonical Divisor on a Fano Variety, *Izv. Akad. Nauk SSSR Ser. Mat.* **43** (1979), English transl.: *Math. USSR Izv.* **14** (1980), pp. 395–405.
[19] Shokurov V. V., The Existence of Lines on Fano 3-Folds, *Izv. Akad. Nauk SSSR Ser. Mat.* **43** (1979), English transl.: *Math. USSR Izv.* **15** (1980), pp. 173–209.
[20] Siu Y. T. and Yau S.-T., Compact Kaehler Manifolds of Positive Bisectional Curvature, *Inv. Math.* **59** (1980), pp. 189–204.
[21] Tsunoda S., Degenerations of Minimal Surfaces with Non-Negative Kodaira Dimension (in Japanese), *Proc. Symp. Algebraic Geometry*, Kinosaki (1981).

NAGOYA UNIVERSITY
NAGOYA, 464, JAPAN

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

ARTHUR OGUS

Periods of Integrals in Characteristic p

One of the most important tools in algebraic geometry over the field $\boldsymbol{C}$ of complex numbers is the theory of periods of integrals, i.e. of the Hodge structures canonically attached to the cohomology of complex algebraic varieties. The yoga of "weights" [4], due to Deligne and Grothendieck, establishes a partial analogy between these Hodge structures and the action of Frobenius (i.e. of $\mathrm{Gal}(\bar{\boldsymbol{F}}_q/\boldsymbol{F}_q)$) on the étale cohomology of a smooth proper scheme over $\boldsymbol{F}_q$. Over the last several years a similar analogy, between the Hodge structures arising from varieties in characteristic 0 and the F-crystals attached to varieties in characteristic p, has begun to emerge. Indeed, because of the connection between crystalline cohomology and differential forms, this analogy often manifests itself as a direct link, which can be quite powerful. In this talk I will describe some of the results and conjectures which currently comprise this analogy, with especial reference to abelian varieties and K3 surfaces as examples.

I will begin by briefly reviewing the basic "outputs" of the machinery of crystalline cohomology ([1], [2]). Let us fix the following notation: k is a perfect field of characteristic $p > 0$, $W = W(k)$ is its Witt ring, and $K(k)$ is the fraction field of $W(k)$. It will be convenient to choose an algebraic closure $\bar{K}$ of $K(k)$, and to let K_{nr} denote the maximal unramified extension of $K(k)$ in $\bar{K}$ and W_{nr} its ring of integers. Finally, K' will denote a finite totally ramified extension of $K(k)$, with ring of integers V.

If X/k is a smooth proper k-scheme, its crystalline cohomology groups $H^i_{\mathrm{cris}}(X/W)$ are finitely generated W-modules, endowed with canonical cup-product maps: $H^i \times H^j \to H^{i+j}$. When X/k is projective or liftable, it is also known that each $H^i_{\mathrm{cris}}(X/W)$ has the expected rank, i.e. $\dim_{\boldsymbol{Q}_l} H^i_{\mathrm{et}}(X \times_k \bar{k}, \boldsymbol{Q}_l)$ $(l \neq p)$. If X has pure dimension N, there is a trace map $\mathrm{tr}\colon H^{2N}(X/W) \to W$, and tr together with cup-product induces the expected Poincaré duality theorem. Crystalline cohomology relates to De Rham cohomology through a long exact universal coefficient sequence:

(1.1) $\to H^i_{\mathrm{cris}}(X/W) \xrightarrow{p} H^i_{\mathrm{cris}}(X/W) \to \boldsymbol{H}^i_{\mathrm{DR}}(X/k) \to H^{i+1}_{\mathrm{cris}}(X/W) \to \ldots$

Moreover, if X has a smooth proper lifting $\underline{X}$ to W, there is a canonical isomorphism:

(1.2) $\sigma_{\mathrm{cris}}\colon \boldsymbol{H}^i_{\mathrm{DR}}(\underline{X}/W) \xrightarrow{\sim} H^i_{\mathrm{cris}}(X/W)$.

Somewhat more generally, if $\underline{X}$ is a lifting of X to V, there is a canonical isomorphism [3]:

(1.3) $\sigma_{\mathrm{cris}}\colon \boldsymbol{H}^i_{\mathrm{DR}}(\underline{X}_{K'}/K') \xrightarrow{\sim} K' \otimes_W H^i_{\mathrm{cris}}(X/W)$.

(1.4) *Remark*. If the absolute ramification index e of V is strictly less than p, it is known that σ_{cris} in fact comes from a map: $\boldsymbol{H}^i_{\mathrm{DR}}(\underline{X}/V) \xrightarrow{\sim} V \otimes H^i_{\mathrm{cris}}(X/W)$. Counterexamples to this "integral" statement are known when e is strictly greater than p, but not when $e = p$ [3]. It would be interesting to construct examples, in the highly ramified case, in which the torsion subgroups of $\boldsymbol{H}^i_{\mathrm{DR}}(\underline{X}/V)$ and of $V \otimes H^i_{\mathrm{cris}}(X/W)$ are not isomorphic.

The absolute Frobenius endomorphism F_X of X induces an endomorphism Φ of $H^i_{\mathrm{cris}}(X/W)$, semi-linear with respect to the Frobenius endomorphism F_W of W. The crystalline analogue of an abstract Hodge structure is the following:

(1.5) Definition. An *F-crystal over k* is an injective F_W-linear endomorphism Φ of a finitely generated W-module H.

It is common to abuse notation by writing H instead of Φ for the F-crystal.

I now want to describe the theorem, due to Mazur [9], [2], which was my original motivation for feeling that F-crystals should be analogous to Hodge structures. Let us assume that (H, Φ) is a torsion free F-crystal over k, and define:

(2.1) $M^iH = \{x \in H\colon \Phi x \in p^iH\}$,

a descending filtration on H. Then Mazur's theorem asserts:

(2.2) Theorem. *Assume that the Hodge spectral sequence of X/k degenerates at E_1 and that the groups $H^*_{\mathrm{cris}}(X/W)$ are all torsion free. Then the image of the natural map: $M^iH^*_{\mathrm{cris}} \to \boldsymbol{H}^*_{\mathrm{DR}}(X/k)$ is precisely the i-th level F^i of the Hodge filtration.*

(2.3) *Remarks.* In fact, Mazur's theorem is somewhat stronger, and suggests that the modules M^i themselves have an important meaning (cf. [2], § 8). It would be very interesting to find geometric illustrations of this deeper structure, which was unfortunately neglected in [14]. Two other natural problems come to mind:

(1) describe what happens when the hypotheses of (2.2) are violated, and

(2) find a version of Mazur's theorem with coefficients in an F-crystal on X (in the sense of [14]).

(2.4) *Example.* Mazur's theorem gives some important restrictions on the nature of the F-crystals which can arise as the cohomology of a specific type of variety. For example, if A/k is an abelian variety of dimension g, $H^1_{\text{cris}}(A/W)$ is free of rank $2g$, and $\Phi\otimes\text{id}_k$ has rank g.

In fact, crystalline cohomology provides the F-crystal $\big(H^1_{\text{cris}}(A/W),\ \Phi\big)$ with an additional bit of structure which we should mention, coming from the trace map. Namely, we let Tr be the composite of the natural isomorphisms [1]:

$$\bigwedge^{2g} H^1_{\text{cris}}(A/W)\to H^{2g}_{\text{cris}}(A/W) \quad \text{and} \quad \text{tr}_{A/W}\colon H^{2g}_{\text{cris}}(A/W)\to W.$$

The F-crystal structure on H^1 induces one on $\bigwedge^{2g} H^1$ in an obvious way and we have the compatibility:

$$(2.4.1)\ \ \text{Tr}\big(\Phi(x)\big) = p^g F^*_W\ \text{Tr}(x) \quad \text{for} \quad x\in \bigwedge^{2g} H^1_{\text{cris}}.$$

We shall call a triple (H,Φ,Tr) satisfying all these conditions an *abelian crystal of genus g.*

Similarly, if X/k is a K3 surface, one finds that $H^2_{\text{cris}}(X/W)$ is a free W-module of rank 22, and that $\Phi\otimes\text{id}_k$ has rank one. Moreover, cup-product and trace induce a nondegenerate symmetric bilinear form:

$$(2.4.2)\ \ (\ |\ )\ H^2_{\text{cris}}\times H^2_{\text{cris}}\to W, \quad \text{satisfying:} \quad (\Phi x\mid \Phi y) = p^2F^*_W(x\mid y).$$

There is an additional, more subtle, condition satisfied by this form: any $x\in\bigwedge^{22} H^2_{\text{cris}}$ such that $\Phi(x) = p^{22}x$ satisfies $(x\mid x) = -1$. For a conjectural generalization of this p-adic analogue of the Hodge index formula and a discussion of some of the added subtleties when $p=2$, I refer the reader to [17] and [15].

Let us now turn to the "concrete" applications of crystalline cohomology to geometry. As expected by Grothendieck, crystalline cohomology has been especially useful as a tool for elucidating the p-adic properties of the geometry of varieties in characteristic p, e.g. the p-adic completion of the Neron–Severi group of a surface or of the endomorphism ring of an abelian variety. The key here is the following very simple result [16]:

(3.1) PROPOSITION. *Suppose k is algebraically closed or finite.*[1]

(3.1.1) *If A/k is an abelian variety, the cokernel of the natural map:*

$$\boldsymbol{Z}_p \otimes \mathrm{End}(A) \to \mathrm{End}\big(H^1_{\mathrm{cris}}(A/W), \Phi\big)$$

is torsion free.

(3.1.2) *If X/k is smooth and proper, if its Hodge spectral sequence degenerates at E_1 and if $H^2_{\mathrm{cris}}(X/W)$ is torsion free, then the cokernel of the Chern class map:*

$$\boldsymbol{Z}_p \otimes \mathrm{NS}(X) \to H^2_{\mathrm{cris}}(X/W)^{\Phi=p}$$

is torsion free.

Remark. I feel sure that more can be said about the Chern class map appearing in (3.1.2); for instance, it may be that the hypothesis on the degeneracy of the Hodge spectral sequence is superfluous. Note that W. Lang and N. Nygaard [8] have in fact used a variant of (3.1.2), involving the De Rham–Witt complex, to *prove* the degeneracy of this spectral sequence for K3 surfaces, thus obtaining a simplified proof of the Rudakov–Shafarevich theorem that in fact $H^0(X, \Omega^1_{X/k}) = 0$.

If k is a finite field, (3.1) has a much stronger conjectural form, which would be an analogue of the Hodge–Lefschetz theorem determining the Neron–Severi group of a complex variety as $F^1_{\mathrm{Hodge}} H^2_{\boldsymbol{Z}}$. Of course, I am referring to the crystalline version of Tate's conjecture.

(3.2) CONJECTURE.

(3.2.1) *If $A/\boldsymbol{F}_q$ is an abelian variety, the map:*

$$\boldsymbol{Z}_p \otimes \mathrm{End}(A/\boldsymbol{F}_q) \to \mathrm{End}\big(H^1_{\mathrm{cris}}(A/W), \Phi\big)$$

is an isomorphism.

(3.2.2) *If $X/\boldsymbol{F}_q$ satisfies the hypotheses of* (3.1.2), *the map:*

$$\boldsymbol{Z}_p \otimes \mathrm{NS}(X) \to H^2_{\mathrm{cris}}(X/W)^{\Phi=p}$$

is an isomorphism.

---

[1] O. Gabber pointed out that (3.1) is valid if k is finite.

Of course, (3.2.1) is in fact a theorem, for it follows easily from Tate's l-adic result [20] and (3.1.1) Moreover, (3.2.2) "tensored with $\boldsymbol{Q}$" should be true for any smooth proper variety. For an application to a refined study of zeta functions of surfaces over finite fields, we refer to work of Milne [10].

The p-adic part of the geometry of varieties which are supersingular (e.g. in the sense that their cohomology is spanned by algebraic cycles) seems to play an especially important role. Recall that an abelian variety $A/\bar{k}$ is supersingular iff $\boldsymbol{Q}\otimes\mathrm{End}(A/\bar{k})$ has rank $4g^2$ over $\boldsymbol{Q}$, and that an abelian crystal (H, Φ, Tr) is supersingular iff $\boldsymbol{Q}\otimes H$ admits a basis of vectors x with $\Phi^2 x = p\cdot x$.

Analogously, a K3-surface $X/\bar{k}$ is supersingular iff $\mathrm{NS}(X)$ has rank 22, and a K3-crystal $\big(H, \Phi, (\ |\ )\big)$ is supersingular iff $\boldsymbol{Q}\otimes H$ admits a basis of vectors x such that $\Phi x = px$ for all x. For such varieties one has the following "Torelli theorem" [16], [17].

(4.1) Theorem. *Let k be an algebraically closed field of characteristic $p > 0$.*

(4.1.1) *The functor H^1_{cris} establishes a bijection between the set of isomorphism classes of supersingular abelian varieties of dimension $g \geqslant 2$ and the set of isomorphism classes of supersingular abelian crystals of genus $g \geqslant 2$.*

(4.1.2) *The functor H^2_{cris} establishes a bijection between the set of isomorphism classes of supersingular* K3 *surfaces and the set of isomorphism classes of supersingular* K3 *crystals (at least if $p \geqslant 5$).*

Of course, the analogue of (4.1.1) in Hodge theory is a tautology, and the analogue of (4.1.2) is a celebrated result of Piatetski–Shapiro and Shafarevich [9].

One of the deepest aspects of Hodge theory is the relationship between families of algebraic varieties and variations of Hodge structures. There are similar links between families of algebraic varieties in characterictic p and families of F-crystals, cf. for instance [14]. In fact, these links played a key role in the proof of (4.1.2), and there is indeed a very precise version of (4.1.2) "in a family". Here let us content ourselves by recalling the following crystalline versions of "local Torelli".

(5.1) Theorem. *Suppose that the absolute ramification index e of V is less than $p-1$. Then:*

(5.1.1) *If A is an abelian variety over k, there is a natural bijection* *from the set of isomorphism classes of formal liftings* $\underline{A}$ *of* A *to* V *to the set of liftings of the Hodge filtration* $F_H^1 H_{\mathrm{DR}}^1(A/k)$ *to* $V \otimes H_{\mathrm{cris}}^1(A/W)$.

Furthermore, a morphism $\eta\colon A \to A'$ *lifts to a morphism* $\underline{A} \to \underline{A}'$ *iff* $H_{\mathrm{cris}}^1(\eta)$ *preserves the corresponding filtrations.*

(5.1.2) *If* X *is a* K3 *surface over* k, *there is a natural bijection from the set of isomorphism classes of formal liftings* $\underline{X}$ *of* X *to* V *to the set of isotropic liftings* Fil *of* $F_H^2 H_{\mathrm{DR}}^2(X/k)$ *to* $V \otimes H_{\mathrm{cris}}^2(X/W)$. *Furthermore, an isomorphism* $\alpha\colon X \to X'$ *lifts to an isomorphism:* $\underline{X} \to \underline{X}'$ *iff* $H_{\mathrm{cris}}^2(\alpha)$ *preserves the filtrations, and a line bundle* L *on* X *lifts to* $\underline{X}$ *iff* $c_{\mathrm{cris}}(L) \in \mathrm{Fil}^{\perp}$.

For applications of these well-known results to the problem of finding algebraizable liftings over DVR's with controlled ramification, we refer to [11] and [16]. It is also possible to classify formal deformations of abelian varieties and K3 surfaces in equicharacteristic p in terms of deformations of the corresponding crystals. However, the precise statement of this result requires some technical preparation which would be out of place here; cf. [13].

Deligne's theory of absolute Hodge cycles [5] is the basis of another link between crystalline cohomology and Hodge theory. In particular, if X is a smooth proper K'-scheme, its group of absolute Hodge cycles $H_{\mathrm{AH}}^*(X)$ is a finite dimensional $\boldsymbol{Q}$-vector space, and there are natural injections:

$$H_{\mathrm{AH}}^*(X) \hookrightarrow \boldsymbol{H}_{\mathrm{DR}}(X/K'), \qquad K' \otimes H_{\mathrm{AH}}^*(X) \to \boldsymbol{H}_{\mathrm{DR}}(X/K').$$

If X has good reduction $\underline{H}/V$, we can use σ_{cris} to identify $\boldsymbol{H}_{\mathrm{DR}}^*(X/K')$ with $K' \otimes_W H_{\mathrm{cris}}^*(\underline{X}_0/W)$. Then one has:

(6.1) CONJECTURE. *If* $z \in H_{\mathrm{AH}}^{2i}(X)$, *its image in* $K' \otimes H_{\mathrm{cris}}^{2i}(X_0/W)$ *lies in* $K(k) \otimes H_{\mathrm{cris}}^{2i}(\underline{X}_0/W)$, *and* $\Phi(z) = p^i z$.

For some partial results, we refer to [15] and [18]. In particular for an abelian variety over a number field (6.1) is true for almost all p. Moreover, if the abelian variety has complex multiplication, the converse is true: a De Rham cohomology class ξ is in the image of H_{AH}^{2i} iff it satisfies $\Phi(\xi) = p^i \xi$ for almost all p.

To continue our parallel between K3 surfaces and abelian varieties, we should prove the Tate conjecture (3.2) for K3 surfaces over finite fields. For "ordinary" K3 surfaces, this was done by N. Nygaard, using the theory of p-adic étale cohomology and p-divisible groups [12]. Recently, he and I have been able to combine these methods with crystalline techniques to obtain a considerably more general result. To state it, we must first recall that F-crystals over an algebraically closed field k can be classified up to isogeny by their slopes, or, equivalently, Newton polygons. If the F-crystal is defined over a finite field $\boldsymbol{F}_q$, for example, these slopes can be computed as follows. If $\boldsymbol{F}_q$ has degree d over the prime field, $F^d_{W_q}$ is the identity map, and hence Φ^d is a W_q-linear endomorphism of a free $\boldsymbol{W}_q$-module. The slopes of the F-crystal are then just the p-adic valuations of the eigenvalues of Φ^d. (For the general definition we refer to [7].)

(7.1) THEOREM (with N. Nygaard). *Suppose X is a* K3 *surface over a finite field $\boldsymbol{F}_q$, and suppose that the smallest slope λ of $H =: \big(H^2_{\mathrm{cris}}(X/W), \Phi\big)$ is less than* 1 *and that $p \geqslant 5$. Then X satisfies Tate's conjecture* (3.2.2).

Tate's conjecture predicts that X should be supersingular (i.e., have $\mathrm{rk}(\mathrm{NS}) = 22$) iff $\lambda = 1$, so this case is excluded from our theorem.

I would now like to provide a sketch of the proof of (7.1), which combines many of the techniques discussed above. Details will appear in [13]. The main idea of the proof is to find a lifting of X which generalizes the "canonical lifting" in the ordinary case $\lambda = 0$. For simplicity, we assume here that $p \geqslant 13$.

To carry out the proof, it will be convenient to make use of the crystalline Weil group [3]:

$$W_{\mathrm{cris}}(\bar{K}) = \{\varphi \in \mathrm{Gal}\big(\bar{K}/K(\boldsymbol{F}_p)\big) \colon \varphi \,|\, W = F^d_W \text{ for some } d \in \boldsymbol{Z}\}.$$

There is an exact sequence: $1 \to \mathrm{Gal}(\bar{K}/\boldsymbol{K}) \to W_{\mathrm{cris}}(\bar{K}) \xrightarrow{\deg} \boldsymbol{Z} \to 1$, where $\boldsymbol{K} = K_{\mathrm{nr}}$ is the fraction field of W and $\deg(\varphi)$ is the integer d appearing in the definition of $W_{\mathrm{cris}}(\bar{K})$. There is a natural semilinear action ϱ of $W_{\mathrm{cris}}(\bar{K})$ on $\bar{K} \otimes H^2_{\mathrm{cris}}(X/W)$, with $\varphi \in W_{\mathrm{cris}}(\bar{K})$ acting by $\varrho_\varphi =: \varphi \otimes \Phi^{\deg(\varphi)}$. Note that in fact ϱ_φ preserves the "lattice" $\bar{V} \otimes H^2_{\mathrm{cris}}(X/W)$, where $\bar{V}$ is the ring of integers of $\bar{K}$.

(7.2) CLAIM. *The number $h =: 1/(1-\lambda)$ is an integer less than or equal to* 10. *There exist: a finite extension K' of degree $\leqslant h$ of $\boldsymbol{K}$, a subgroup Γ of finite index in $W_{\mathrm{cris}}(\bar{K})$, and an isotropic lifting* Fil *of $F^2_{\mathrm{Hodge}} H^2_{\mathrm{DR}}(X/\boldsymbol{F})$ to $V \otimes H^2_{\mathrm{cris}}(X/W)$ which is stable under Γ. Moreover, if $\gamma \in \Gamma$ and $\omega \in \mathrm{Fil}$, $\varrho_\gamma(\omega) = \chi_\gamma \omega$ where $\mathrm{ord}_p(\chi_\gamma) = (2-\lambda)\deg(\gamma)$.*

The proof is just semi-linear algebra, thanks to Mazur's determination (2.1) of the Hodge filtration of X in terms of the endomorphism Φ.

(7.3) Claim. *If* $\xi \in H^2_{\mathrm{cris}}(X/W)$ *and* $\Phi(\xi) = p\xi$, *then* $1 \otimes \xi \in \mathrm{Fil}^{\perp}$.

This claim is an immediate consequence of the compatibility of the action of W_{cris} with the pairing $(\ |\ )$. Indeed, if $\gamma \in W_{\mathrm{cris}}$ and $\omega \in \mathrm{Fil}$, we have:

$$p^{2\deg(\gamma)}\gamma(\omega \mid \xi) = (\varrho_\gamma \omega \mid \varrho_\gamma \xi) = p^{\deg(\gamma)}(\omega \mid \xi).$$

Since $\lambda < 1$, $\mathrm{ord}_p(\chi_\gamma) > p^{\deg(\gamma)}$, and since we can choose γ with $\deg(\gamma) \neq 0$, $(\omega \mid \xi) = 0$.

(7.4) Claim. *There is an algebraic lifting* $\underline{X}$ *of* X *to* V *such that* $\sigma_{\mathrm{cris}}\left(\mathrm{Fil}^2_{\mathrm{Hodge}}\, H^2_{\mathrm{DR}}(\underline{X}/V)\right) = \mathrm{Fil}$. *Every line bundle on* $\underline{X}$ *lifts to* X.

It follows immediately from (5.1.2) that Fil determines a formal lifting $\underline{X}$ of X to V, and every line bundle on X will lift to $\underline{X}$ by (7.3) and (5.1.2). By Grothendieck's existence theorem, $\underline{X}$ is algebraizable.

For the next step of the proof, we choose an isomorphism $\sigma\colon \overline{K} \to \boldsymbol{C}$, which we use to obtain a K3 surface σX over $\boldsymbol{C}$. We let $H^i_\sigma(X)$ denote the singular cohomology of σX with $\boldsymbol{Q}$ coefficients. Assuming without loss of generality that Γ fixes K', we see that if $\varphi \in \Gamma$,

$$\overline{K} \underset{\varphi}{\otimes} H^2_{\mathrm{DR}}(\underline{X}/\overline{K}) \cong \overline{K} \underset{\varphi}{\otimes} H_{\mathrm{DR}}(\underline{X}/K') \cong \overline{K} \otimes H_{\mathrm{DR}}(\underline{X}/K'),$$

and so ϱ_φ also induces a $\overline{K}$-linear endomorphism $\tilde{\varrho}_\varphi$ of $H^2_{\mathrm{DR}}(\underline{X}/\overline{K})$. The isomorphism σ allows us to view $\tilde{\varrho}_\varphi$ as an endomorphism of $H^2(\sigma\underline{X}/\boldsymbol{C}) \cong \boldsymbol{C} \underset{\boldsymbol{Q}}{\otimes} H^2_\sigma(\underline{X})$.

(7.5) Claim. *After* Γ *is replaced by a subgroup of finite index, it will have the following property: for every* $\varphi \in \Gamma$, *the endomorphism* $\tilde{\varrho}_\varphi$ *of* $\boldsymbol{C} \otimes H_\sigma(\underline{X})$ *is a morphism of Hodge structures, i.e. it preserves the Hodge filtration and the* $\boldsymbol{Q}$*-subspace* $H_\sigma(X)$.

It is clear from the construction that $\tilde{\varrho}_\varphi$ will preserve the Hodge filtration. The proof that it preserves $H_\sigma(X)$ is more difficult. In [18] it is shown that there exists an abelian variety A/K and an absolute Hodge cycle u mapping the primitive cohomology of $\underline{X}/\overline{K}$ into $\mathrm{End}\, H^1(A)$. Furthermore, A has good reduction and u satisfies conjecture (6.1), i.e., is compatible with the actions of $W_{\mathrm{cris}}(\overline{K})$. A calculation, based on the construction of A and (7.2), shows that Γ preserves the Hodge filtration of A. Then a variant of (5.1.1) implies that Γ acts on A via isogenies, and hence preserves $H^1_\sigma(A)$. It follows that Γ also preserves $H^2_\sigma(\underline{X})$.

To prove (7.1), let K' be a field of definition of $\underline{X}$, with residue field $\boldsymbol{F}_{p^{d'}}$. It will suffice to prove that if $d =: d'm$ is sufficiently large, the rank of $\mathrm{NS}(\underline{X})$ is at least as big as the dimension t of the p^d-eigenspace E of Φ^d acting on $H^2_{\mathrm{cris}}(X/W(\boldsymbol{F}_{p^d}))$. If $\varphi \in \Gamma$ has degree d and is contained in $\mathrm{Gal}(\overline{K}/K')$ the linear action $\tilde{\varrho}_\varphi$ on $H^2_{\mathrm{DR}}(\underline{X}/\overline{K}) \cong \overline{K}\otimes H^2_{\mathrm{cris}}(X/W(\boldsymbol{F}_{p^{d'}}))$ is just $\mathrm{id}_{\overline{K}}\otimes\Phi^d$. Since this action preserves $H^2_\sigma(\underline{X})$, its p^d-eigenspace is defined over $H^2_\sigma(\underline{X})$, i.e. $E\cap H^2_\sigma(\underline{X})$ has $\boldsymbol{Q}$-rank t. By (7.3), $E \subseteq \mathrm{Fil}^{\perp} = \mathrm{Fil}^1_{\mathrm{Hodge}}H^2(\underline{X})$, and so the Lefschetz–Hodge theorem implies that $t = \mathrm{rk}\,\mathrm{NS}(\underline{X})$. □

I would like to finish by formulating a conjecture which is inspired by the above proof and which is, in a way, a complement to (6.1). Let $Y/\overline{K}$ be a smooth and proper scheme with good reduction, so that $W_{\mathrm{cris}}(\overline{K})$ acts on $H^*_{\mathrm{DR}}(Y/\overline{K})$, as above, and hence also on any tensor construction $T_{\mathrm{DR}}(Y/\overline{K})$ applied to De Rham cohomology. We define two subgroups of $W_{\mathrm{cris}}(\overline{K})$ as follows:

Γ_{H}: $\{\gamma \in W_{\mathrm{cris}}$: ϱ_γ preserves the Hodge filtration of $T_{\mathrm{DR}}(Y/\overline{K})\}$,

Γ_{AH}: $\{\gamma \in W_{\mathrm{cris}}$: ϱ_γ is absolutely Hodge$\}$.

(8.1) CONJECTURE. *The inclusion $\Gamma_{\mathrm{AH}} \subseteq \Gamma_{\mathrm{H}}$ is in fact an equality.*

It is easy to see that this conjecture is true for abelian varieties, if $T = H^1$. In [13], we shall prove it for K3 surfaces (satisfying a mild technical condition) when $T = H^2$.

(8.2) DEFINITION. A variety $Y/\overline{K}$ as above is *quasi-canonical* (*with respect* to T) iff Γ_{AH} is of finite index in W_{cris}.

With this terminology, we can say that the key point in the proof of (7.1) was the construction of a quasi-canonical lifting of our K3 surface to characteristic zero, with the aid of (8.1). For varieties defined over a local field of mixed characteristic, one can define an l-adic version of Γ_{AH} in terms of the action of the (usual) Weil group on étale cohomology. Here one sees an advantage of the crystalline theory: its link with differential forms made conjecture (8.1) natural, and its concomitant link with deformation theory then allowed us to construct the quasi-canonical lifting.

References

[1] Berthelot P., *Cohomologie cristalline des schémas de caractéristique $p > 0$*, Lecture Notes in Math. **407**, Berlin-Heidelberg-New York, Springer, 1974.

[2] Berthelot P. and Ogus A., Notes on crystalline cohomology, *Math. Notes* **21**, Princeton University Press, 1978.

[3] Berthelot P. and Ogus A., *F*-isocrystals and De Rham cohomology I, *Inv. Math.*, **72** (1983), pp. 159–199.
[4] Deligne P., Poids dans la cohomologie des variétés algébriques, *Proc. Int. Cong.* 1974, pp. 79–86, Canadian Math. Congress (1975)
[5] Deligne P., Hodge cycles on abelian varieties (notes by J. S. Milne), in: *Hodge cycles, motives nad Shimura varieties*, Lecture Notes in Math. **900**, Berlin-Heidelberg-New York, Springer, 1962.
[6] Deligne P., Illusie, L., Cristaux ordinaires et coordonnées canoniques, in: *Surfaces algébriques*, Lecture Notes in Math. **868**, Springer, 1981.
[7] Katz N., Slope filtration of *F*-crystals, *Asterisque* **63** (1979), pp. 113–144.
[8] Lang W. and Nygaard N., A short proof of the Rudakov–Shafarevich theorem, *Math. Ann.* (1980), pp. 171–173.
[9] Mazur B., Frobenius and the Hodge filtration-estimates, *Ann. of Math.* (1973), pp. 58–95.
[10] Milne J., On a conjecture of Artin and Tate, *Ann. of Math.* **102** (1975), pp. 517–533.
[11] Norman P., Lifting abelian varieties, *Inv. Math.* **64**, (1981), pp. 431-444.
[12] Nygaard N., The Tate conjecture for ordinary K3 surfaces over finite fieds, *Inv. Math.* **74** (1983), pp. 213–237.
[13] Nygaard N., Ogus A., *The Tate conjecture for more* K3 *surfaces of finite height*, in preparation.
[14] Ogus A., *F*-crystals and Griffiths transversality, *Proc. Int. Symp. Alg. Geom. Kyoto* **1977**, Kinokuniya (1978).
[15] Ogus A., Hodge cycles and crystalline cohomology. In: *Hodge cycles, motives and Shimura varieties*, Lect. Notes in Math. **900**, Springer (1982).
[16] Ogus A., Supersingular K3 crystals, *Asterisque* **64** (1979), pp. 3–86.
[17] Ogus A., *A crystalline Torelli theorem for supersingular* K3 *surfaces*, to appear.
[18] Ogus A., *F-isocrystals and De Rham cohomology II — convergent isocrystals*, to appear in *Duke Journal of Math.*
[19] Piatetski-Shapiro I. and Shafarevich I., A Torelli theorem for surfaces of type K3, *Math. U.S.S.R. Izvestiya* **5** (1971), pp. 545–588.
[20] Tate J., Endomorphism of abelian varieties over finite fields, *Inv. Math* **2** (1966), pp. 134–144.

UNIVERSITY OF CALIFORNIA
BERKELEY, CA.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

BERNARD TEISSIER

Sur la classification des singularités des espaces analytiques complexes

Introduction

On associe à chaque point x d'un espace analytique complexe réduit X la suite des multiplicités en x des *variétés polaires locales* de X. Étant donné un plongement $(X, x) \subset (\boldsymbol{C}^N, 0)$, on peut décrire la variété polaire locale de codimension k soit comme ensemble des points $y \in X$ où une direction limite en y d'hyperplans tangents à la partie lisse X^0 de X appartient à un sous-espace linéaire $\boldsymbol{P}^{d-k} \subset \check{\boldsymbol{P}}^{N-1}$ donné et assez général, soit comme lieu critique d'une projection linéaire assez générale $X \to \boldsymbol{C}^{d-k+1}$. Dans le cas où X est un cône de sommet x on retrouve des notions connues de géométrie projective. La première définition permet de relier les variétés polaires à la géométrie des limites d'espaces tangents et en particulier de montrer l'existence d'une stratification de Whitney canonique minimale de X, c'est-à-dire de classifier, du point de vue de l'équisingularité à la Whitney, les singularités de X. La seconde définition permet de relier, au moyen de la théorie de Morse, les multiplicités des variétés polaires en x à des invariants topologiques locaux de X en x. Ceci permet de démontrer une réciproque du théorème de Thom–Mather, c'est-à-dire de montrer l'équivalence de conditions topologiques et des conditions de Whitney. Une partie de ces résultats s'étend au cas relatif, c'est-à-dire à la condition w_f de Thom stricte pour un morphisme f: $X \to S$. Les liens avec la théorie des $\mathscr{D}$-modules d'une part, et avec les problèmes de résolution simultanée des singularités d'autre part, auraient par trop allongé le texte. Le lecteur est renvoyé à [10], [17] pour l'un et à [13] pour l'autre.

§ 1. Limites d'espaces tangents

Soit N un entier et soit $X \subset \boldsymbol{R}^N$ un sous-ensemble sous-analytique. D'après [6], il existe un sous-ensemble sous-analytique $\operatorname{Sing} X$ de X

tel que $X^0 = X - \mathrm{Sing}\, X$ soit dense dans X et soit un sous-espace analytique réel non singulier de $\boldsymbol{R}^N$. Supposons que toutes les composantes connexes de X^0 aient la même dimension d, et considérons l'application de Gauss $\gamma\colon X^0 \to G(N-1, d-1)(\boldsymbol{R})$ de X^0 dans la grassmanienne des $(d-1)$-plans de $\boldsymbol{P}^{N-1}(\boldsymbol{R})$ qui à chaque point $x \in X^0$ associe la direction de l'espace tangent à X^0 en x. En utilisant le théorème de rectilinéarisation de *loc. cit.*, on peut vérifier que le graphe de l'application de Gauss est sous-analytique, et il en est donc de même (*loc. cit.*) de son adhérence $N(X)$ dans $X \times G$ où $G = G(N-1, d-1)(\boldsymbol{R})$. Le sous-ensemble sous-analytique $N(X)$ de $\boldsymbol{R}^N \times G$, muni du morphisme $\nu\colon N(X) \to X$ induit par la première projection, sera appelé *modification de Nash* de l'ensemble sous-analytique X. Le morphisme ν induit un isomorphisme $\nu^{-1}(X^0) \xrightarrow{\sim} X^0$ et $\dim N(X) = d$. Pour chaque point $x \in X$, le sous-ensemble sous-analytique $\nu^{-1}(x) = N(X) \cap (\{x\} \times G)$ de G est l'ensemble des positions limites en x d'espaces tangents aux points de X^0. De façon analogue, on peut définir l'espace conormal de X dans $\boldsymbol{R}^N$ comme l'adhérence $T^*_X\boldsymbol{R}^N$, dans l'espace cotangent $T^*\boldsymbol{R}^N$ à $\boldsymbol{R}^N$, de l'espace conormal à X^0 dans $\boldsymbol{R}^N$ constitué de l'ensemble des couples $(x, \xi) \in T^*\boldsymbol{R}^N$ tels que x appartienne à X^0 et que l'espace tangent à X^0 en x soit contenu dans le noyau de l'application linéaire $\xi\colon T_{\boldsymbol{R}^N,x} \to \boldsymbol{R}$. Ainsi $T^*_X\boldsymbol{R}^N$ est un sous-ensemble fermé de $T^*\boldsymbol{R}^N \simeq \boldsymbol{R}^N \times \check{\boldsymbol{R}}^N$ où $\check{\boldsymbol{R}}^N$ désigne le dual de $\boldsymbol{R}^N$, stable par les homothéties du second facteur et dont on vérifie comme plus haut qu'il est sous-analytique. On note $C(X, \boldsymbol{R}^N)$ son projectivisé, c'est-à-dire son image dans $\boldsymbol{R}^N \times \check{\boldsymbol{P}}^{N-1}$, et $\varkappa\colon C(X, \boldsymbol{R}^N) \to X$ le morphisme induit par la première projection. Un point (x, ξ) de $T^*\boldsymbol{R}^N$ est dans $T^*_X\boldsymbol{R}^N$ si ξ est limite en x de directions d'hyperplans tangents en des points de X^0, c'est-à-dire s'il existe une suite de points $x_i \in X^0$ tendant vers x et d'hyperplans $\xi_i = 0$ tangents à X^0 en x_i, c'est-à-dire tels que T_{X^0,x_i} soit contenu dans le noyau de ξ_i, et que l'on a $\xi = \mathrm{Lim}\, \xi_i$. Il revient au même de dire que le noyau de ξ en x contient une direction limite en x d'espaces tangents à X^0. On remarque que pour $x \in X^0$, la fibre $\varkappa^{-1}(x)$ est isomorphe à $\boldsymbol{P}^{N-1-d}$, et donc $\dim C(X, \boldsymbol{R}^N) = N-1$.

Soit $f\colon X \to \boldsymbol{R}^p$ un morphisme sous-analytique, que nous supposons installé, c'est-à-dire que l'on s'est donné un plongement $X \subset \boldsymbol{R}^p \times \boldsymbol{R}^N$ sous-analytique et que f est obtenu comme restriction à X de la première projection. Supposons qu'il existe un ouvert sous-analytique dense X^0 de X qui soit une sous-variété analytique réelle de X telle que en chaque point x de X^0 le noyau de l'application linéaire tangente à f en x soit de dimension d. On peut définir une application de Gauss relative $\gamma_f\colon X^0 \to G$ et par l'adhérence de son graphe définir comme plus haut la modification de Nash relative, notée $\nu_f\colon N_f(X)\to X$; un point $z\in N_f(X)$ représente une direction limite en $\nu_f(z)$ d'espaces tangents aux fibres de f en des points de X^0 tendant vers $\nu_f(z)$. De même on peut définir l'espace conormal relatif $T^*_{X^0/\boldsymbol{R}^p}\boldsymbol{R}^p\times\boldsymbol{R}^N$ de X^0 dans $\boldsymbol{R}^p\times\boldsymbol{R}^N$ comme noyau de l'homomorphisme naturel de fibrés cotangents relatifs $T^*(\boldsymbol{R}^p\times\times\boldsymbol{R}^N)/\boldsymbol{R}^p\underset{\boldsymbol{R}^p\times\boldsymbol{R}^N}{\times}X^0\to T^*(X^0/\boldsymbol{R}^p)$ et l'adhérence dans $T^*(\boldsymbol{R}^p\times\boldsymbol{R}^N)/\boldsymbol{R}^p$ de $T^*_{X^0/\boldsymbol{R}^p}(\boldsymbol{R}^p\times\boldsymbol{R}^N)/\boldsymbol{R}^p$ est l'espace conormal relatif de X, qui est un sous-ensemble sous-analytique de $T^*(\boldsymbol{R}^p\times\boldsymbol{R}^N)/\boldsymbol{R}^p\simeq\boldsymbol{R}^p\times\boldsymbol{R}^N\times\check{\boldsymbol{R}}^N$ stable par rapport aux homothéties de $\check{\boldsymbol{R}}^N$, et dont le projectivisé est un sous-ensemble sous-analytique $C_f(X,\boldsymbol{R}^p\times\boldsymbol{R}^N)$ de $\boldsymbol{R}^p\times\boldsymbol{R}^N\times\check{\boldsymbol{P}}^{N-1}(\boldsymbol{R})$, de dimension $N+p-1$, en fait contenu dans $X\times\check{\boldsymbol{P}}^{N-1}(\boldsymbol{R})$.

Supposons maintenant que X est un espace analytique complexe réduit et soit $f\colon X\to S$ un morphisme analytique complexe. Notons $\Omega^1_f=\Omega^1_{X/S}$ le $\mathcal{O}_X$-module cohérent des 1-formes différentielles relatives et supposons qu'il existe un sous-ensemble ouvert dense X^0 de X tel que $\Omega^1_f\mid X^0$ soit localement libre de rang d. Soit $g\colon G\to X$ la grassmannienne des quotients localement libres de Ω^1_f. Par définition, l'image réciproque $g^*\Omega^1_f$ a sur G un quotient localement libre de rang d: $g^*\Omega^1_f\to L\to 0$ et, pour un morphisme $h\colon T\to X$, se donner un quotient localement libre de rang d, $h^*\Omega^1_f\to L_T\to 0$, équivaut à se donner un morphisme $T\xrightarrow{s}G$ tel que $g\circ s=h$ et $s^*L\simeq L_T$. D'après l'hypothèse on a donc une section $\sigma\colon X^0\to G$ de $g\mid g^{-1}(X^0)$. L'adhérence de l'image de cette section est un sous-ensemble analytique fermé $N_f(X)\subset G$, et le morphisme $\nu_f\colon N_f(X)\to X$ induit par g est propre et induit un isomorphisme au-dessus de l'ouvert dense $\nu_f^{-1}(X^0)$. C'est donc une modification de X, appelée modification de Nash relative de X/S. Si le morphisme f est installé au moyen d'un plongement $X\subset S\times\times\boldsymbol{C}^N$, ce qui est toujours le cas localement sur X, on retrouve ainsi la modification de Nash décrite ci-dessus dans le cadre sous-analytique. On peut de même, dans le cas où f est installée, définir l'espace conormal relatif (complexe) de X dans $S\times\boldsymbol{C}^N$, qui est un sous-espace analytique de $T^*(S\times\boldsymbol{C}^N)/S\simeq S\times\boldsymbol{C}^N\times\check{\boldsymbol{C}}^N$ conique par rapport aux homothéties de $\check{\boldsymbol{C}}^N$, et l'espace conormal projectif associé $C_f(X,S\times\boldsymbol{C}^N)$ ou $C_f(X)$, qui est un sous-espace analytique fermé de $X\times\check{\boldsymbol{P}}^{N-1}$, de dimension $N-1+\dim S$.

Exemples. (1) Soit U un ouvert de $\boldsymbol{C}^N$ et soit $f\colon U\to\boldsymbol{C}$ un morphisme analytique. La modification de Nash relative coïncide avec l'espace conormal relatif et est l'éclatement dans U de l'idéal jacobien $j(f)=(\partial f/\partial z_1,\ldots,\partial f/\partial z_N)\mathcal{O}_U$ engendré par les dérivées partielles de f.

(2) Soient U un ouvert de $\boldsymbol{C}^N$ et $X = f^{-1}(0)$ avec f comme ci-dessus et tel que $f^{-1}(0)$ soit réduit non vide. La modification de Nash de X coïncide avec l'espace conormal et est l'éclatement dans X de l'idéal $j(f)\cdot\mathcal{O}_X$ engendré dans $\mathcal{O}_X$ par les restrictions à X des dérivées partielles de f (voir [26], [27]).

Soit maintenant X un sous-espace analytique complexe réduit, purement de dimension d, d'un ouvert U de $\boldsymbol{C}^N$, et soit x un point de X. Considérons le diagramme:

$$\begin{tikzcd}
X\times \boldsymbol{P}^{N-1}\times\check{\boldsymbol{P}}^{N-1} & E_xC(X) \arrow[l, hook'] \arrow[r, "\tilde{e}_x"] \arrow[d, "\varkappa'"'] \arrow[dr, "\zeta"'] & C(X) \arrow[r, hook] \arrow[d, "\varkappa"] & X\times\check{\boldsymbol{P}}^{N-1} \\
X\times \boldsymbol{P}^{N-1} & E_xX \arrow[l, hook'] \arrow[r, "e_x"'] & X \arrow[r, hook] & \boldsymbol{C}^N
\end{tikzcd}$$

où $\varkappa$ est le morphisme conormal défini ci-dessus, e_x l'éclatement du point x dans X, $\tilde{e}_x$ l'éclatement dans $C(X)$ du sous-espace $\varkappa^{-1}(x)$, et $\varkappa'$ le morphisme dû à la propriété universelle de l'éclatement. Soit $\mathscr{Y} = \bigcup_a \mathscr{Y}_a$ la réunion des composantes irréductibles de dimension $N-2$ du diviseur exceptionnel $\zeta^{-1}(x)$, que nous considérons comme sous-espace de $\boldsymbol{P}^{N-1}\times\times\check{\boldsymbol{P}}^{N-1}$. Pour chaque a posons $V_a = \varkappa'(|\mathscr{Y}_a|) \subset \boldsymbol{P}^{N-1}$, et $W_a = \tilde{e}_x(|\mathscr{Y}_a|) \subset \check{\boldsymbol{P}}^{N-1}$, où $|\mathscr{Y}_a|$ désigne l'espace réduit sous-jacent à $\mathscr{Y}_a$. Remarquons que puisque $|\mathscr{Y}| = |\zeta^{-1}(x)|$ d'après le Hauptidealsatz, on a $\bigcup_a V_a = |e_x^{-1}(x)| = |\mathrm{Proj}\ C_{X,x}|$ où $C_{X,x}$ désigne le cône tangent à X en x, et $\bigcup_a W_a = |\varkappa^{-1}(x)|$, ensemble des limites en x d'hyperplans tangents à X.

En fait, la collection des sous-variétés projectives V_a de $|\mathrm{Proj}\ C_{X,x}|$ suffit pour reconstituer l'ensemble des limites en x d'hyperplans tangents:

Théorème (Lê–Teissier). *Pour chaque a, W_a est la variété projective duale de V_a, et par conséquent l'ensemble des limites en x d'hyperplans tangents à X est réunion des variétés duales des V_a. De plus, pour chaque composante irréductible C de $|\mathrm{Proj}\ C_{X,x}|$ il existe exactement un indice a tel que l'on ait $V_a = C$.*

En fait, on montre que pour chaque a, $|\mathscr{Y}_a|$ est l'espace conormal $C(V_a, \boldsymbol{P}^{N-1})$, ce qui implique que W_a est le dual projectif de V_a.

Exemple (cf. [17]). Soit $(X, 0) \subset (\boldsymbol{C}^3, 0)$ un germe de surface analytique complexe réduite. Il existe un nombre fini de génératrices $l_1, \ldots, l_r$ du cône tangent $C_{X,0}$ telles que l'espace $|\nu^{-1}(0)|$ des limites en 0 d'espaces tangents à X^0 soit la réunion du dual de la courbe projective $|\mathrm{Proj}\, C_{X,0}|$ et des droites projectives $\check{l}_i = \{H \in \boldsymbol{C}^3 | H \supset l_i\}$. De plus, si $(X, 0)$ est à singularité isolée, ou si $C_{X,0}$ est réduit, la déformation naturelle de $(X, 0)$ sur son cône tangent $C_{X,0}$ est équisingulière (cf. ci-dessous) si et seulement si $r = 0$.

Remarque. Il est probable qu'un avatar du théorème ci-dessus est vrai dans le cas sous-analytique. Pour l'étude numérique des limites d'espaces tangents, voir ([26], Chap. 2, 1.6), et pour l'étude géométrique, voir [4], [14], [17].

§ 2. Variétés polaires locales

Soient N un entier, et

$$\mathfrak{D}\colon (0) \subset D_{N-1} \subset D_{N-2} \subset \ldots \subset D_1 \subset D_0 = \boldsymbol{C}^N$$

un drapeau de sous-espaces vectoriels de $\boldsymbol{C}^N$, avec $\operatorname{codim} D_i = i$. Étant donnés un entier d, $0 < d \leqslant N$ et une suite d'entiers $a = (a_1, \ldots, a_d)$, on considère dans la grassmanienne $G = G(N-1, d-1)$ des sous-espaces vectoriels de dimension d de $\boldsymbol{C}^N$ la sous-variété algébrique (variété de Schubert)

$$\sigma_a(\mathfrak{D}) = \{T \in G | \dim(T \cap D_{d+a_i-i}) \geqslant i \text{ pour } i = 1, \ldots, d\}.$$

Sa codimension dans G est purement $\sum a_i$. Dans le cas particulier où $a = (1, \ldots, 1, 0, \ldots, 0)$ avec k fois 1, on notera $c_k(\mathfrak{D})$ pour $\sigma_a(\mathfrak{D})$.

Soit maintenant $f\colon X \to S$ un morphisme analytique complexe installé comme au § 1, c'est-à-dire muni d'un S-plongement $X \subset S \times \boldsymbol{C}^N$. Supposons S non singulier, et considérons la modification de Nash relative du § 1

$$\begin{array}{ccc} & & G \\ & \overset{\gamma_f}{\nearrow} & \uparrow \mathrm{pr}_2 \\ N_f(X) & \hookrightarrow & X \times G \\ \downarrow & & \\ X & & \end{array}$$

et le morphisme de Gauss relatif γ_f. À l'aide d'un théorème de Kleiman et de l'ouverture de la transversalité, on peut montrer (cf. [28], Chap. IV, 1.3) que tout point $x \in X$ possède un voisinage ouvert, que nous noterons encore X, tel qu'il existe un ouvert de Zariski dense W de l'espace des drapeaux, tel que pour tout $\mathfrak{D} \in W$, on ait pour tout a:

(i) $\gamma_f^{-1}(\sigma_a(\mathfrak{D})) \cap \nu_f^{-1}(X^0)$ est réduit et dense dans $|\gamma_f^{-1}(\sigma_a(\mathfrak{D}))|$, et ce dernier espace est vide ou purement de codimension $\sum a_i$ dans $N_f(X)$.

(ii) Si $\gamma_f^{-1}(\sigma_a(\mathfrak{D})) \cap \nu_f^{-1}(x)$ n'est pas vide, on a

$$\dim \gamma_f^{-1}(\sigma_a(\mathfrak{D})) \cap \nu_f^{-1}(x) = \dim \nu_f^{-1}(x) - \sum a_i.$$

En conséquence, pourvu que le drapeau $\mathfrak{D}$ soit assez général, si l'on pose

$$P_a\langle f; \mathfrak{D}\rangle = \left|\nu_f\big(\gamma_f^{-1}(\sigma_a(\mathfrak{D}))\big)\right|,$$

on voit que $P_a\langle f; \mathfrak{D}\rangle$ est l'adhérence dans X de $P_a\langle f; \mathfrak{D}\rangle \cap X^0$, et est un sous-espace analytique réduit purement de codimension $\sum a_i$ de X ou vide. De plus, $\gamma_f^{-1}(\sigma_a(\mathfrak{D}))$ coïncide ensemblistement avec le transformé strict de $P_a\langle f; \mathfrak{D}\rangle$ par le morphisme ν_f (noter que $\gamma_f^{-1}(\sigma_a(\mathfrak{D}))$ n'est pas en général réduit).

DÉFINITION. $P_a\langle f; \mathfrak{D}\rangle$ sera appelée *variété polaire locale* associée à f: $X \to S$, au drapeau $\mathfrak{D}$, et à la suite $a = (a_1, \ldots, a_d)$.

PROPOSITION. *Dans cette situation, pour chaque point $x \in X$ et chaque a, il existe un ouvert de Zariski dense W_x de l'espace des drapeaux tel que la multiplicité $m_x(P_a\langle f; \mathfrak{D}\rangle)$ de $P_a\langle f; \mathfrak{D}\rangle$ en x ne dépende pas de $\mathfrak{D} \in W_x$, et en fait ce nombre ne dépend que du type analytique du germe en x du morphisme f.*

Lorsque $a = (\overbrace{1, \ldots, 1}^{k}, 0, \ldots, 0)$, on notera $P_k\langle f; \mathfrak{D}\rangle$ pour $P_a\langle f; \mathfrak{D}\rangle$. C'est un sous-espace analytique réduit, purement de codimension k dans X ou vide. Lorsque S est un point, on notera $P_k\langle X; \mathfrak{D}\rangle$ pour $P_k\langle f; \mathfrak{D}\rangle$. On note que $P_k\langle f; \mathfrak{D}\rangle$ ne dépend en fait que de D_{d-k+1}, car l'on a $c_k(\mathfrak{D}) = \{T \in G \mid \dim T \cap D_{d-k+1} \geqslant k\}$. On le notera aussi $P_k\langle f; D_{d-k+1}\rangle$.

On propose ici d'associer à un morphisme f: $X \to S$, en chaque point x de X, les invariants analytiques $m_x(P_a\langle f; \mathfrak{D}\rangle)$ où $\mathfrak{D}$ est assez général. On s'intéressera particulièrement aux invariants $m_x(P_k\langle f; \mathfrak{D}\rangle)$, $k = 0, \ldots$ $\ldots, d$. Dans le cas particulier où S est un point et X un espace analytique réduit, on associera, pour tout point $x \in X$, au germe (X, x) (ou à l'algèbre $\mathcal{O}_{X,x}$) la suite de d entiers

$$M^*_{X,x} = \big(m_x(P_0\langle X; \mathfrak{D}\rangle), m_x(P_1\langle X; \mathfrak{D}\rangle), \ldots, m_x(P_{d-1}\langle X; \mathfrak{D}\rangle)\big).$$

On remarquera que $P_0\langle X; \mathfrak{D}\rangle = X$ et par conséquent le premier terme de cette suite est la multiplicité de X en x.

Exemples. (1) Reprenons l'exemple (1) ci-dessus. On a $d = N-1$ et la modification de Nash relative est contenue dans le sous-espace de $U \times \check{P}^{N-1}$ défini par les équations $T_i(\partial f/\partial z_j) - T_j(\partial f/\partial z_i) = 0$. Dans $\check{P}^{N-1}$ les seules variétés de Schubert sont les $c_k(\mathfrak{D})$, et en choisissant des coordonnées adaptées au drapeau $\mathfrak{D}$, on voit que $P_k\langle f; \mathfrak{D}\rangle$ est obtenu comme ceci. Soit $\operatorname{Sing} f$ le lieu critique de f, et pour chaque entier k, soit $\tilde{P}_k$ le sous-espace de U défini par $\partial f/\partial z_1 = \ldots = \partial f/\partial z_k = 0$. Alors $P_k\langle f; \mathfrak{D}\rangle$ est l'adhérence dans U de $\tilde{P}_k \setminus \operatorname{Sing} f$. Dans le cas où f a un seul point critique dans U, noté $0, P_k\langle f; \mathfrak{D}\rangle$ est exactement le sous-espace de U défini par $\partial f/\partial z_1 = \ldots = \partial f/\partial z_k = 0$. On retrouve la définition des variétés polaires donnée dans ([27], [14]). Dans ce cas (cf. [27]), la multiplicité en 0 de $P_k\langle f; \mathfrak{D}\rangle$ (toujours pour $\mathfrak{D}$ assez général) est égale à $\mu^{(k)}(f, 0)$, qui est par définition le nombre de Milnor ([23]) de la restriction de f à un sous-espace vectoriel de dimension k assez général de C^N. De plus, le nombre d'intersection en 0 de la courbe polaire $P_{N-1}\langle f; \mathfrak{D}\rangle$ avec l'hypersurface $f^{-1}(f(0))$ vaut $\mu^{(N)}(f, 0) + \mu^{(N-1)}(f, 0)$.

(2) Reprenons l'exemple (2) ci-dessus en supposant X réduit. Puisque les valeurs critiques de f sont isolées, on peut vérifier que l'on a $P_k\langle X; \mathfrak{D}\rangle = P_k\langle f; \mathfrak{D}\rangle \cap X$. Si 0 est un point critique isolé de f, on a les égalités:

$$m_0(P_k\langle X; \mathfrak{D}\rangle) = \mu^{(k)}(f, 0) + \mu^{(k+1)}(f, 0) \qquad (0 \leqslant k \leqslant N-2).$$

(3) Soit $V \subset P^{N-1}$ une variété projective réduite et soit $X \subset C^N$ le cône de sommet 0 sur V. Alors les variétés polaires $P_k\langle X; \mathfrak{D}\rangle$ sont les cônes sur les lieux polaires M_k de V introduits par Todd dans le cas où V est non singulière et par R. Piene [24] dans le cas général, leurs multiplicités sont les classes de V.

La théorie classique des variétés polaires projectives correspond donc au cas particulier où S est un point et X un cône de sommet 0. Par exemple, si X est le cône sur une courbe projective plane réduite de degré m et de classe ($=$ degré de la courbe duale) $\check{m}$, on a $M^*_{X,0} = (m, \check{m})$.

Les variétés polaires "absolues" ($S =$ un point) sont introduites et étudiées dans [18]. On a dans [23, chap. IV] la description suivante des variétés polaires relatives comme lieux critiques de projections. Supposons donné un S-plongement $X \subset S \times C^N$, et supposons que f soit un morphisme lisse en tout point $x \in X^0$. Soit p: $C^N \to C^{d-k+1}$ une projection linéaire. Pour $x \in X^0$, la fibre $X(f(x))$ est non singulière en x, contenue dans $\{f(x)\} \times C^N$ et l'on notera π_x: $X(f(x)) \to C^{d-k+1}$ la restriction à la fibre $X(f(x))$ de la projection p. Soit $P_k\langle f; p\rangle^0$ l'ensemble des points $x \in X^0$ tels que x soit critique pour π_x. Alors on a $P_k\langle f; p\rangle^0 = \nu_f\big(\gamma_f^{-1}(c_k(D_{d-k+1})) \cap \nu_f^{-1}(X^0)\big)$, et par conséquent pour p assez générale, l'adhérence $P_k\langle f; p\rangle$ de $P_k\langle f; p\rangle^0$ dans X est la variété polaire locale $P_k\langle f; \operatorname{Ker} p\rangle$.

On doit à J. P. G. Henry et M. Merle (cf. [5]) la description suivante de $P_k\langle f; \mathfrak{D}\rangle$. Supposons toujours $X \subset S \times C^N$ et considérons l'espace conormal relatif du § 1:

$$\begin{tikzcd} & \check{P}^{N-1} \\ C_f(X) \arrow[ur, "\lambda_f"] \arrow[r, hook] \arrow[d] & X \times \check{P}^{N-1} \arrow[u, "\mathrm{pr}_2"'] \\ X & \end{tikzcd}$$

Alors pour tout $x \in X$ et $0 \leqslant k \leqslant d = \dim X - \dim S$, il existe un voisinage ouvert de x dans X, encore noté X, et un ouvert de Zariski dense U_k de la grassmanienne des $(d-k)$-plans de $\check{P}^{N-1}$ tels que pour $L^{d-k} \in U_k$, on ait $\left|\varkappa_f\left(\lambda_f^{-1}(L^{d-k})\right)\right| = P_k\langle X; D_{d-k+1}\rangle$, où D_{d-k+1} est l'intersection des hyperplans appartenant à L^{d-k}.

En fait, la considération des cônes tangents des variétés polaires permet de reconstruire géométriquement les sous-variétés projectives V_α de $|\operatorname{Proj} C_{X,x}|$ vues au § 1, et donc les limites d'espaces tangents à X en x:

Théorème (Lê–Teissier). *Étant données* $(X,0) \subset (C^N, 0)$ *comme ci-dessus et une projection linéaire* $p\colon C^N \to C^{d-k+1}$ *assez générale, le réduit* $|C_{P_k,0}|$ *du cône tangent en* 0 *à la variété polaire* $P_k = P_k\langle X; \operatorname{Ker} p\rangle$ *est la réunion des cônes sur les variétés projectives* V_α *du* § 1 *qui vérifient* $\dim V_\alpha + 1 = d-k$ (partie fixe) *et des variétés polaires de dimension* $d-k$ *relatives à* $\operatorname{Ker} p$ *des cônes sur les* V_α *telles que* $\dim V_\alpha + 1 > d-k$ (partie mobile).

§ 3. Conditions de Whitney et de Thom

Étant donnés deux sous-espaces vectoriels A et B de C^N, muni de la métrique hermitienne, on définit la distance de A à B (dans cet ordre) par la formule

$$\delta(A, B) = \sup_{\substack{u \in B^\perp - \{0\} \\ v \in A - \{0\}}} \frac{|(u, v)|}{\|u\| \cdot \|v\|},$$

où $B^\perp = \{u \in C^N \mid (u, b) = 0 \text{ pour tout } b \in B\}$. On note que $\delta(A, B) = 0$ si et seulement si $B \supset A$.

Soient $f\colon X \to S$ un morphisme entre espaces analytiques réduits, tel qu'il existe un ouvert dense X^0 de X sur lequel f est de corang constant d. Soit Y un sous-espace analytique non singulier de X sur lequel f est de corang constant, et x un point de Y. On dit que le couple (X^0, Y) satisfait la *condition* a_f *de Thom* en x s'il existe un plongement local $(X, x) \subset (\boldsymbol{C}^M, 0)$ tel que, pour toute suite $(x_n)_{n\in N}$ de points de X^0 tendant vers x, on a

$$\operatorname*{Lim}_{n\to\infty} \delta\Big(T_x Y\big(f(x)\big),\ T_{x_n} X\big(f(x_n)\big)\Big) = 0$$

ou: toute limite en x d'espaces tangents aux fibres de $f \mid X^0$ contient l'espace tangent en x à la fibre de $f \mid Y$.

On dit que (X^0, Y) satisfait la *condition* a_f *de Thom stricte avec exposant* 1, aussi appelée *condition* w_f s'il existe un voisinage U de x dans X et une constante $C > 0$ tels que pour tout $y \in U \cap Y$ et $x' \in U \cap X^0$ on ait

$$\delta\Big(T_y Y\big(f(y)\big),\ T_{x'} X\big(f(x')\big)\Big) \leqslant C\ \operatorname{dist}(x', Y).$$

Dans le cas où S est un point, on dira que le couple (X^0, Y) satisfait la *condition* (a) *de Whitney* (resp. la *condition* (w), ou condition (a) de Whitney stricte au sens de Hironaka avec exposant 1).

Cette condition est indépendante du choix du plongement.

Restons dans le cas où S est un point, et choisissons une rétraction locale $\varrho\colon \boldsymbol{C}^M \to Y$. On dira que (X^0, Y) satisfait en $x \in Y$ la *condition* (b) *de Whitney* (resp. la *condition* (b) *stricte avec exposant* e) si pour toute suite $(x_n)_{n\in N}$ de points de X^0 tendant vers x, on a $\operatorname*{Lim}_{n\to\infty}\operatorname{dist}\big(\widehat{x_n\varrho(x_n)},\ T_{x_n}X^0\big) = 0$ (resp. s'il existe un voisinage U de x et une constante C tels que l'on ait $\operatorname{dist}\big(\widehat{x\varrho(x)},\ T_x X^0\big) \leqslant C \operatorname{dist}(x, Y)^e$ pour $x \in X^0$) où $\widehat{x\varrho(x)}$ désigne la direction de la droite sécante joignant x et $\varrho(x)$ dans $\boldsymbol{C}^M$.

On appelle stratification d'un espace analytique X une partition localement finie $X = \bigcup X_\alpha$ de X en sous-espaces analytiques non singuliers à fermeture et frontière analytiques.

Définition. Une stratification de X est une *stratification de Whitney* si pour tout couple X_α, X_β tel que $X_\alpha \cap \bar{X}_\beta \neq \emptyset$ on a $X_\alpha \subset \bar{X}_\beta$ et en tout $x \in X_\alpha$, le couple (X_β, X_α) satisfait les conditions (a) et (b) de Whitney.

On appelle stratification d'un morphisme $f\colon X \to S$ la donnée d'une stratification $X = \bigcup X_\alpha$ de X, et d'une stratification $S = \bigcup S_\beta$ de S telles que la restriction de f à chaque X_α soit une submersion de X_α sur une strate S_β. Ainsi chaque $f^{-1}(S_\beta)$ est réunion de strates X_α.

Définition (Thom–Sabbah, cf. [25]). On dit qu'un morphisme stratifié f est *sans éclatement* si

(1) Pour tout couple de strates X_α, X_β de X telles que $X_\alpha \cap \bar{X}_\beta \neq \emptyset$ on a:

$X_\alpha \subset \bar{X}_\beta$ et (X_β, X_α) satisfait la condition $a_{f|\bar{X}_\beta}$ en tout point de X_α.

(2) La stratification $S = \bigcup S_\beta$ de S est une stratification de Whitney, et enfin:

(3) Pour tout β, la stratification $f^{-1}(S_\beta) = \bigcup\limits_{X_\alpha \subset f^{-1}(S_\beta)} X_\alpha$ est de Whitney.

Les stratifications de Whitney et de Thom–Sabbah ont des propriétés opologiques extrêmement utiles:

Théorème (Thom–Mather, cf. [21]). *Si $X = \bigcup X_\alpha$ est une stratification de Whitney, pour tout $x \in X$, tout plongement local $(X, x) \subset (\mathbf{C}^N, 0)$ il existe une rétraction locale r: $(\mathbf{C}^N, 0) \to (X_\alpha, x)$, où X_α est la strate contenant x, et un nombre réel $\varepsilon_0 > 0$ tels que pour tout $0 < \varepsilon < \varepsilon_0$, il existe $\eta_\varepsilon > 0$ tel que pour $0 < \eta < \eta_\varepsilon$ on ait, en notant $\boldsymbol{B}_\varepsilon$ la boule de centre 0 et de rayon ε dans $\mathbf{C}^N$, un homéomorphisme $\boldsymbol{B}_\varepsilon \cap r^{-1}(\boldsymbol{B}_\eta \cap X_\alpha) \simeq \big(r^{-1}(x) \cap \boldsymbol{B}_\varepsilon\big) \times (X_\alpha \cap \boldsymbol{B}_\eta)$ compatible avec la rétraction r et induisant pour chaque $\bar{X}_\beta$ contenant x un homéomorphisme de*

$$\bar{X}_\beta \cap \boldsymbol{B}_\varepsilon \cap r^{-1}(\boldsymbol{B}_\eta \cap X_\alpha) \quad \textit{sur} \quad \big(\bar{X}_\beta \cap r^{-1}(x)\big) \times (X_\alpha \cap \boldsymbol{B}_\eta).$$

En d'autres termes, chaque $\bar{X}_\beta$, ou si on préfère, l'ensemble stratifié X est localement topologiquement trivial le long de X_α en x.

Théorème (Thom–Lê–Sabbah). *Si f: $X \to S$ est un morphisme stratifié sans éclatement, f admet une théorie des cycles évanescents en tout point $x \in X$.* (Voir [25].)

Rappelons que cela signifie que pour tout $x \in X$, posant $s = f(x)$, il existe un système fondamental à un paramètre de bons voisinages de x (resp. s) dans X (resp. S), noté $(B_\varepsilon)_{\varepsilon \in I}$ (resp. $(D_\eta)_{\eta \in I'}$) où $I =]0, a]$, $I' =]0, a']$ (cf. [19], § 2), et un sous-ensemble sous-analytique E de $I \times I'$ auquel $I \times 0$ est adhérent tels que pour $(\varepsilon, \eta) \in E$, le morphisme induit par f:

$$f_{\varepsilon,\eta}\colon B_\varepsilon \cap f^{-1}(D_\eta) \to D_\eta$$

est descriptible, c'est-à-dire qu'il existe une stratification de Whitney analytique complexe de D_η, en fait $D_\eta = \bigcup (D_\eta \cap S_\beta)$, telle que $f_{\varepsilon,\eta}$ induise une fibration topologique localement triviale $f_{\varepsilon,\eta}^{-1}(D_\eta \cap S_\beta) \to D_\eta \cap S_\beta$ pour tout β. Ceci implique en particulier que les faisceaux images directes

$R^i(f_{\varepsilon,\eta})_* C$ sont analytiquement constructibles sur D_η et que l'on a de plus, pour $\varepsilon' < \varepsilon$ et η' tel que $(\varepsilon', \eta') \in E$, l'égalité $R^i(f_{\varepsilon',\eta'})_* C = R^i(f_{\varepsilon,\eta})_* C \mid D_{\eta'}$.

En outre, les stratifications de Whitney et de Thom–Sabbah ont le mérite d'exister!

THÉORÈME (Whitney [32]). *Soient X un espace analytique complexe (réduit) et $(Y_i)_{i\in I}$ une famille localement finie de sous-espaces analytiques à fermeture analytique dans X. Il existe une stratification de Whitney $X = \bigcup X_\alpha$ telle que chaque Y_i soit réunion de strates.*

Rappelons que l'existence de stratifications de Whitney pour les ensembles semi-(resp. sous-) analytiques a été démontrée par Łojasiewicz [20] (resp. Hironaka [8] et Hardt [34]).

THÉORÈME (Sabbah [25]). *Soit $f: X \to S$ un morphisme propre d'espaces analytiques complexes réduits muni d'une stratification $\mathscr{S}$. Il existe un éclatement π: $S' \to S$ de S et une stratification $\mathscr{S}'$ du produit fibré $X' = X \times_S S'$, compatible avec $\mathscr{S}$ en ce sens que les images réciproques dans S' (resp. X') des strates de S (resp. X) sont réunions de strates, et telle que le morphisme f': $X' \to S'$ déduit de f par changement de base soit sans éclatement. Dans le cas où f est complexifié d'un morphisme analytique réel, l'éclatement est aussi complexifié d'un morphisme analytique réel, et les strates des stratifications sont invariantes par conjugaison.*

En particulier, si S est une courbe non singulière, f admet une théorie des cycles évanescents.

Remarque. Sabbah a également montré dans *loc. cit.* l'analogue local du Théorème ci-dessus.

Exemple. Reprenons l'exemple (1) ci-dessus et soit Y un sous-espace non singulier de U contenu dans l'hypersurface $f^{-1}(0)$. Supposons $0 \in Y$ et $f(0) = 0$. Choisissons des coordonnées locales $y_1, \ldots, y_t, z_1, \ldots, z_{N-t}$ pour C^N en 0 telles que Y soit défini par: $z_1 = \ldots = z_{N-t} = 0$. Dans [29], j'ai introduit la condition d'équisingularité appelée condition (C) pour X le long de Y en 0:

$$\left(\frac{\partial f}{\partial y_i}\right)_0 \in \overline{(z_1, \ldots, z_{N-t})\left(\frac{\partial f}{\partial z_1}, \ldots, \frac{\partial f}{\partial z_{N-t}}\right)\mathcal{O}_{C^N,0}} \qquad (1 \leqslant i \leqslant t),$$

où la barre désigne la fermeture intégrale des idéaux. Dans [27] on montre que la condition (C) implique les conditions (a) et (b) de Whitney et dans [2] la réciproque, dans le cas où $X^0 = X - Y$. Dans [27] (appendice) on montre que la condition (C) équivaut à la condition w_f pour (X^0, Y) en 0, et d'autre part un exemple de Briançon–Speder, $f = z_1^3 + yz_1z_2^3 + z_2^4z_3 + z_3^9$, satisfait a_f en 0 mais pas w_f.

Par ailleurs, Kuo–Verdier (cf. [31]) ont montré que la condition (w) implique les conditions (a) et (b) de Whitney, dans le cadre sous-analytique réel.

§ 4. Critères numériques pour les conditions de Whitney et de Thom

THÉORÈME (cf. [28], chap. V, [22]). *Soient X un espace analytique complexe réduit purement de dimension d, Y un sous-espace analytique de X et $0 \in Y$ un point non singulier de Y. Les conditions suivantes sont équivalentes:*

(1) *Le couple (X^0, Y) satisfait les conditions* (a) *et* (b) *de Whitney en* 0.

(2) *L'application $Y \to \mathbf{N}^d$ définie par $y \mapsto M^*_{X,y}$ est constante au voisinage de* 0.

(3) *Le couple (X^0, Y) satisfait la condition* (w) *en* 0, *donc au voisinage de* 0.

La démonstration est par récurrence sur $d - t$ où $t = \dim_0 Y$. On utilise de façon fondamentale le diagramme analogue à celui du § 1, où e_Y est l'éclatement de Y:

$$\begin{tikzcd}
 & & & \check{\mathbf{P}}^{N-1} \\
 & E_Y C(X; \mathbf{C}^N) \arrow[r, "\tilde{e}_Y"] \arrow[d, "\varkappa'_Y"'] \arrow[dr, "\zeta"] & C(X; \mathbf{C}^N) \arrow[ur] \arrow[d, "\varkappa"] & \subset X \times \check{\mathbf{P}}^{N-1} \arrow[u, "\mathrm{pr}_2"'] \\
X \times \mathbf{P}^{N-1-t} \supset & E_Y X \arrow[r, "e_Y"'] & X \subset \mathbf{C}^N &
\end{tikzcd}$$

qui permet de relier les cônes normaux le long de Y des variétés polaires aux limites d'hyperplans tangents. Une étape importante de la démonstration de (2) $\Rightarrow$ (3) est de prouver que (2) est équivalent à

(2′) *Le morphisme* ζ: $\zeta^{-1}(Y) \to Y$ *a toutes ses fibres de la même dimension au voisinage de* 0.

On doit à J. P. G. Henry et M. Merle la découverte du fait que l'équivalence de (2) et (2′) nécessite la considération du diagramme ci-dessus à la place du diagramme analogue construit à partir de la modification de Nash. On peut supposer choisi un plongement local de X dans C^N tel que Y soit un sous-espace linéaire de C^N. L'idée de la récurrence est de prouver que chacune des conditions implique que si H est un hyperplan de C^N assez général parmi ceux contenant Y, alors $|X \cap H|$ satisfait les mêmes conditions. À part des résultats généraux sur l'équimultiplicité et une idée de Hironaka prouvant (2′)$\Rightarrow$(3), les résultats utilisées sont des généralisations et renforcements de l'idée de "transversalité des variétés polaires aux plans servant à les définir" dont le prototype est le résultat suivant (cf. [19]): si $\mathfrak{D}$ est un drapeau assez général, pour chaque i, l'espace vectoriel D_i est transverse en 0 dans C^N à la variété polaire $P_{d-i}\langle X; D_{i-1}\rangle$ en ce sens que l'on a $|D_i \cap C_0(P_{d-i}\langle X; D_{i-1}\rangle)| = \{0\}$, où $C_0(P_{d-i}\langle X; D_{i-1}\rangle)$ désigne le cône tangent en 0 à la variété polaire.

Pour les courbes polaires, on a un résultat beaucoup plus précis inspiré par un résultat de [1] dans le cas $N = 3$, qui est très utile pour la preuve de (1)$\Rightarrow$(2). Soit $(X, 0) \subset (C^N, 0)$, et soit $(Y, 0) \subset (X, 0)$ une courbe non singulière. Pour une projection linéaire p: $C^N \to C^2$ assez générale, notant $P_{d-1}\langle X; \operatorname{Ker} p\rangle$ la courbe polaire associée à la projection p, *la même* projection p est une projection générale pour la courbe $Y \cup P_{d-1}\langle X; \operatorname{Ker} p\rangle$ en ce sens qu'aucune direction limite en 0 de bisécantes à cette courbe n'est contenue dans $\operatorname{Ker} p$ (cf. [28], chap. V, lemme clé).

Remarque. On doit pouvoir démontrer le résultat analogue en géométrie analytique réelle en traduisant géométriquement le raisonnement algébrique de *loc. cit.*

Comme on a ainsi traduit les conditions de Whitney en termes purement algébriques (dimension des fibres de ζ et multiplicités de variétés polaires) ne dépendant que des algèbres locales $\mathcal{O}_{X,x}$ (en fait de leur complété $\hat{\mathcal{O}}_{X,x}$) on peut utiliser les résultats généraux sur la semi-continuité de la multiplicité de la manière suivante (cf. [18], [28], chap. VI).

Soient X un espace analytique, $(Y_i)_{i \in I}$ une famille localement finie de sous-espaces analytiques fermés de X. Définissons une filtration de X inductivement comme suit: $F_0 = X$ et supposons avoir défini $F_0, \ldots, F_{k-1}$. Pour chaque i, notons $F_i = \bigcup F_{i,j_i}$ la décomposition en composantes irréductibles de F_i. Définissons F_k comme la réunion du lieu singulier de F_{k-1}, et de la famille localement finie de sous-espaces analytiques fermés rares dans F_{k-1} que voici: les $F_{k-1} \cap Y_j$ rares dans F_{k-1} et les $\{x \in F_{k-1,j_{k-1}} |$ l'une des suites $M^*_{X,x}$, $M^*_{F_{1,j_1},x}$, $\ldots$, $M^*_{F_{k-2,j_{k-2}},x}$, ne prend pas en x la valeur qu'elle prend en un point général de $F_{k-1,j_{k-1}}\}$.

La filtration $X = F_0 \supset F_1 \supset \ldots F_{k-1} \supset F_k \supset \ldots$ stationne localement, et l'on a:

PROPOSITION. *La famille des composantes connexes des $F_k - F_{k+1}$ est une stratification de Whitney de X telle que chaque Y_i soit réunion de strates. Toute stratification de Whitney ayant ces propriétés en est un raffinement, c'est-à-dire que chaque F_k est réunion de strates de cette stratification* (cf. [28], chap. VI, [19]).

En particulier, prenant $I = \emptyset$, on a:

COROLLAIRE. *Pour tout espace analytique réduit localement équidimensionnel, la construction ci-dessus fournit une stratification de Whitney à strates connexes, dite* canonique, *dont toute stratification de Whitney de X est un raffinement.* On appelle *filtration canonique* de X la filtration construite ci-dessus.

Remarques. La stratification canonique est invariante par le pseudogroupe des isomorphismes locaux de X. Pour tout ouvert U de X, la famille des $F_k \cap U$ est la filtration canonique de U. Enfin pour tout plongement local $(X, x) \subset (C^N, 0)$ et tout sous-espace non singulier $(H, 0)$ de $(C^N, 0)$ transverse dans C^N en 0 à la strate contenant x, il existe un voisinage ouvert U de x dans X tel que les $|F_k \cap H \cap U|$ constituent la filtration canonique de $|U \cap H|$.

COROLLAIRE. *Si X et les Y_i sont des variétés algébriques complexes définies sur une extension galoisienne K de Q, les fermés F_i sont des sous-variétés algébriques de X définies sur K.*

En particulier, comme l'a remarqué Deligne, nous pouvons considérer un germe $(X, 0) \subset (C^N, 0)$ défini par des polynômes $P_j = \sum_A p_{j,A} Z^A$, $1 \leqslant j \leqslant r$. Notons M le nombre des $p_{j,A}$, introduisons des indéterminés $P_{j,A}$, et soit $\mathscr{X}$ la sous-variété de $C^M \times C^N$ définie par les équations $\sum_A P_{j,A} Z^A = 0$ $(1 \leqslant j \leqslant r)$ et la filtration canonique $\mathscr{X} = F_0 \supset F_1 \supset \ldots$ associée à $\mathscr{X}$ et au sous-espace $C^M \times \{0\}$.

Notons π: $\mathscr{X} \to C^M$ la première projection. Quitte à raffiner la stratification associée aux F_k, on peut supposer que pour chaque strate $\mathscr{X}_a$ contenue dans $C^M \times \{0\}$, $\pi^{-1}(\mathscr{X}_a)$ est réunion de strates. D'après le théorème de Thom–Mather, $\pi^{-1}(\mathscr{X}_a)$ est localement topologiquement trivial le long de $\mathscr{X}_a$ et donc pour $p, q \in \mathscr{X}_a$, les germes définis par $\sum p_{j,A} Z^A = 0$ $(1 \leqslant j \leqslant r)$ et par $\sum q_{j,A} Z^A$ $(1 \leqslant j \leqslant r)$ sont topologiquement équivalents (et en fait bien plus, voir ci-dessous). Puisque les $\mathscr{X}_a$ sont définis sur $\mathbf{Q}$, les points à coordonnées algébriques y sont denses, et par conséquent n'importe quel germe algébrique est topologiquement équivalent (et en fait bien plus) avec un germe défini par des polynômes à coefficients entiers algébriques.

Ceci répond en particulier à une question que posait F. Pham en 1972 au séminaire de Thom, et implique que l'ensemble des germes algébriques à (w)-équisingularité près est dénombrable; la classification de ceux-ci est donc possible.

On a, avec une preuve analogue, un analogue relatif du théorème ci-dessus. Soit $f: X \to S$ comme au § 1. Pour tout $x \in X$ notons $P_k(f, x)$ le germe en x de "la" variété polaire générale locale de codimension k, et posons $\Sigma_k(f) = \{x \in X \mid P_k(f, x) \neq \emptyset\}$. (Si f est installé, $\Sigma_k(f) = \{x \in X \mid \dim \varkappa_f^{-1}(x) \geqslant N - d + k\}$.)

DÉFINITION (Sabbah). On dit que f est *sans éclatement en codimension* 0 si pour tout k, $0 \leqslant k \leqslant d$, les fibres de $f \mid \Sigma_k(f): \Sigma_k(f) \to S$ sont de dimension $\leqslant d - k$.

THÉORÈME (Henry–Merle–Sabbah, cf. [5]). *Soient $f: X \to S$ un morphisme sans éclatement en codimension* 0, *Y un sous-espace analytique non singulier d'une fibre $f^{-1}(0)$. Les conditions suivantes sont équivalentes:*

(1) *Le couple (X^0, Y) satisfait la condition w_f en tout point de Y.*

(2) *Les variétés polaires générales $P_k(f)$ $(0 \leqslant k \leqslant d)$ sont équimultiples le long de Y.*

§ 5. Critères topologiques pour les conditions de Whitney

La présentation des variétés polaires locales comme lieux critiques de restrictions à $X \subset \mathbf{C}^N$ de projections linéaires $p: \mathbf{C}^N \to \mathbf{C}^{d-k+1}$ vue au § 2 permet d'appliquer la théorie de Morse pour calculer les multiplicités en x des variétés polaires locales de X en x en fonction d'invariants topologiques locaux de X, généralisant ainsi les formules de l'exemple (2) du § 2. Voici les invariants topologiques locaux qui interviennent (cf. [9], [10], [3], [19]).

Soit $X = \bigcup X_a$ une stratification de Whitney de X. Pour $x \in X_a$ soit $(X, x) \subset (\mathbf{C}^N, 0)$ un plongement local. Pour chaque entier $i \geqslant d_a$, où $d_a = \dim X_a$, il existe un ouvert de Zariski dense W de la grassmanienne des plans de codimension i de $\mathbf{C}^N$ et pour chaque $L_i \in W$ deux nombres réels ε_0, η_0, un sous-ensemble sous-analytique $E \subset]0, \varepsilon_0] \times]0, \eta_0]$ auquel $]0, \varepsilon_0] \times \{0\}$ est adhérent et un sous-ensemble sous-analytique U de $\bigcup_{\eta \in]0,\eta_0]} \boldsymbol{B}_\eta$ tel que chaque $U_\eta = U \cap \boldsymbol{B}_\eta$ soit dense dans $\boldsymbol{B}_\eta$, tels que pour tout $(\varepsilon, \eta) \in E$ et $t \in U_\eta$ la caractéristique d'Euler–Poincaré $\chi\big(X \cap (L_i + t) \cap \boldsymbol{B}_\varepsilon\big)$ ne dépende que de (X, X_a), i.e., pas du plongement, ni du choix de $x \in X_a$ (ici $L_i + t$ est un $(N-i)$-plan parallèle à L_i et passant par t). On la notera $\chi_i(X, X_a)$.

THÉORÈME ([19], 4.11). *Soit $X = \bigcup X_a$ un espace analytique complexe réduit muni d'une stratification de Whitney analytique complexe et soit $x \in X_a$. On a l'égalité*

$$\chi_{d_a+1}(X, X_a) - \chi_{d_a+2}(X, X_a) = \sum_{\beta \neq a} (-1)^{d_\beta - d_a - 1} m_x\big(P_{d_\beta - d_a - 1}(\overline{X}_\beta, x)\big)\big(1 - \chi_{d_\beta+1}(X, X_\beta)\big).$$

On vérifie aussitôt que cette formule permet de calculer par récurrence les $m_x\big(P_k(\overline{X}_\beta, x)\big)$ en fonction des caractéristiques d'Euler–Poincaré $\chi_j(X_\beta, X_\gamma)$, et inversement.

Ce résultat, joint à la caractérisation numérique des stratifications de Whitney vue ci-dessus, a permis dans *loc. cit.* de caractériser les stratifications de Whitney des espaces analytiques complexes par des conditions topologiques: Soit $X = \bigcup X_a$ une stratification analytique complexe. Disons qu'elle satisfait la *condition* (T T)* (= trivialité topologique locale pour les sections générales) si pour toute strate X_a, $x \in X_a$ et tout plongement local $(X, x) \subset (\boldsymbol{C}^N, 0)$, il existe pour tout i, $0 \leqslant i \leqslant N - d_a$ un ouvert de Zariski dense W_i de la grassmanienne des plans de codimension i de $\boldsymbol{C}^N$ contenant $T_{X_a,x}$ tel que pour tout sous-espace non singulier L de $\boldsymbol{C}^N$ défini au voisinage de 0, contenant X_a et tel que $T_{L,0} \in W_i$, l'intersection $\overline{X}_\beta \cap L$ est localement topologiquement triviale le long de X_a en x pour toute strate X_β telle que $x \in \overline{X}_\beta$. On a alors

THÉORÈME (Lê–Teissier, [19]). *Pour une stratification analytique complexe $X = \bigcup X_a$ d'un espace analytique X, les conditions suivantes sont équivalentes:*

(1) $X = \bigcup X_a$ *est une stratification de Whitney.*

(2) $X = \bigcup X_a$ *satisfait la condition* (TT)*.

On a ainsi, dans ce cas, la réciproque correcte du théorème de Thom–Mather. En fait dans *loc. cit.* on associe à chaque point d'un espace analytique complexe son "type d'homotopie évanescent total" qui est un invariant combinatoire local de X et on montre que les conditions ci-dessus sont aussi équivalentes à:

(3) *le type d'homotopie évanescent total de X est constant le long de chaque strate.*

Il se trouve que la formule ci-dessus, appliquée dans le cas où X est le cône sur une variété projective $V \subset \boldsymbol{P}_{N-1}$, contient une généralisation des formules de Plücker, c'est-à-dire permet de calculer le degré de la variété duale $V^{\vee}$ de V (i.e., sa classe) en fonction de la topologie de l'ensemble stratifié que constitue V munie de sa stratification de Whitney canonique. Pour le voir il suffit de faire les observations suivantes:

(1) La classe de V est égale à $m_0(P_{d-1}\langle X; \mathfrak{D}\rangle)$, où $d = \dim X$.

(2) Si $V = G_0 \supset G_1 \supset \ldots \supset G_i \supset \ldots$ est la filtration canonique de V, notant F_i le cône de sommet 0 sur G_i, la filtration canonique de X est $X = F_0 \supset F_1 \supset \ldots \supset \{0\}$ (parfois le $\{0\}$ est superflu).

(3) Si $L+t$ est un hyperplan voisin de 0 dans $\boldsymbol{C}^N$ et général, on a

$$\chi\big(X \cap (L+t) \cap \boldsymbol{B}_\varepsilon\big) = \chi(V) - \chi(V \cap H) \qquad \text{où} \qquad H = \operatorname{Proj} L.$$

(4) Si L_i+t est un plan de codimension i voisin de 0 dans $\boldsymbol{C}^N$, il existe un hyperplan L_1+t voisin de 0 et un plan L_{i-1} de codimension $i-1$ passant par 0 tel que $L_i+t = L_{i-1} \cap (L_1+t)$. On a donc $\chi\big(X \cap (L_i+t) \cap \boldsymbol{B}_\varepsilon\big) = \chi(V \cap H_{i-1}) - \chi(V \cap H_{i-1} \cap H_1)$ où $H_i = \operatorname{Proj} L_i$.

(5) Pour toute strate X_α de X autre que $\{0\}$, notant $V_\alpha = \operatorname{Proj} X_\alpha$, on a

$$\chi_k(X, X_\alpha) = \chi_{k-1}(V, V_\alpha).$$

On retrouve ainsi facilement, dans le cas où V est une hypersurface de degré m à singularités isolées de $\boldsymbol{P}^{d+1}$ la formule de [29] et [12]:

$$\deg V^{\vee} = m(m-1)^d - \sum_{x_i \in \operatorname{sing} V} \big(\mu^{(d+1)}(V, x_i) + \mu^{(d)}(V, x_i)\big).$$

(Comparer à [24], [11], [35].)

Enfin il faut remarquer que l'on n'a pas encore donné de réciproque au théorème de Thom–Lê–Sabbah du § 3 dans un cas plus général que celui de l'exemple du § 3.

Notes ajoutées à la correction des épreuves:

1. J'ai appris à Varsovie que la sous-analycité du morphisme de Gauss du § 1 était prouvé dans [36].

2. Dans l'énoncé du théorème de Thom–Mather au § 3, il existe en fait, par l'analycité, un sous-ensemble sous-analytique E de $]0, a] \times]0, a']$, auquel $]0, a] \times \{0\}$ est adhérent, et tel que pour $(\varepsilon, \eta) \in E$, on ait...

Références

[1] Briançon J. et Henry J. P. G., Equisingularité générique des familles de surfaces à singularités isolées, *Bull. Soc. Math. France* **108** (1980), pp. 259–281.

[2] Briançon J. et Speder J. P., Les conditions de Whitney impliquent «μ^* constant», *Ann. Inst. Fourier* **26** (2) (1976), pp. 153–163.

[3] Dubson A., Classes caractéristiques des variétés singulières, *C. R. Acad. Sci. Paris Sér. A* **287** (1978), pp. 237–240.

[4] Henry J. P. G. et Lê D. T., Limites d'espaces tangents. In: *Séminaire Norguet*, Lecture Notes in Math., 482, Springer, 1975, pp. 251–265.

[5] Henry J. P. G., Merle M. et Sabbah C., Sur la condition de Thom stricte pour un morphisme analytique complexe, *Ann. Sci. École Norm. Sup.* **17** (1984), à paraître.

[6] Hironaka H., *Introduction to Real-Analytic Sets and Real-Analytic Maps*, Istituto Matematico «L. Tonelli», Università di Pisa, Pisa, 1973.

[7] Hironaka H., Stratifications and Flatness. In: *Real and Complex Singularities, Nordic Summer School, Oslo 1976*, Sijthoff and Noordhöff, 1977, pp. 199–265.

[8] Hironaka H., Subanalytic Sets. In: *Number Theory, Algebraic Geometry and Commutative Algebra, Volume in honour of Y. Akizuki*, Kinokuniya, Tokyo, 1973.

[9] Kashiwara M., Index Theorem for a Maximally Overdetermined System of Linear Differential Equations, *Proc. Japan. Acad. Sci.* **49** (1973), pp. 803–804.

[10] Kashiwara M., *Systèmes d'équations microdifférentielles*, cours rédigé par T. Monteiro-Fernandes, Université de Paris-Nord, 1978.

[11] Kleiman S., The Enumerative Theory of Singularities. In: *Real and Complex Singularities, Nordic Summer School, Oslo 1976*, Sijthoff and Noordhoff, 1977, pp. 297–396.

[12] Laumon G., Degré de la variété duale d'une hypersurface à singularités isolées, *Bull. Soc. Math. France* **104** (1976), pp. 51–63.

[13] Laufer H., Weak Simultaneous Resolution for Deformations of Gorenstein Surface Singularities. In: *Singularities*. Proc. Sympos. Pure Math., 40, 1983, part 2, pp. 1–29.

[14] Lê D. T., Calcul du nombre de cycles évanouissants d'une hypersurface complexe, *Ann. Inst. Fourier* **23** (4) (1973), pp. 261–270.

[15] Lê D. T., Some Remarks on Relative Monodromy. In: *Real and Complex Singularities, Nordic Summer School, Oslo 1976*, Sijthoff and Noordhoff, 1977, pp. 397–403.

[16] Lê D. T. et Mebkhout Z., Variétés caractéristiques et variétés polaires, *C. R. Acad. Sci. Paris Sér. I* **296** (1983), pp. 129–132.

[17] Lê D. T. et Teissier B., Sur la géométrie des surfaces complexes I. Tangentes exceptionnelles, *Amer. J. Math.* **101** (1979), pp. 420–452.

[18] Lê D. T. et Teissier B., Variétés polaires locales et classes de Chern des variétés singulières, *Ann. of Math.* (2) **114** (1981), pp. 457–491.

[19] Lê D. T. et Teissier B., Cycles évanescents, sections planes, et conditions de Whitney II. In: *Singularities*, Proc. Sympos. Pure Math., 40, 1983, part 2, pp. 65–103.

[20] Łojasiewicz S., *Ensembles semi-analytiques*, preprint multigraphié, I. H. E. S., Bures-sur-Yvette, 1965.

[21] Mather J., Stratifications and Mappings. In: *Dynamical Systems*, Academic Press, New York, 1973, pp. 195–232.

[22] Merle M., Variétés polaires, stratifications de Whitney et classes de Chern des espaces analytiques complexes (d'après Lê–Teissier), *Séminaire Bourbaki, Nov. 82, Astérisque* **105–106** (1983), pp. 65–78.

[23] Milnor J., *Singular Points of Complex Hypersurfaces*, Ann. of Math. Studies, 61, Princeton University Press, Princeton, 1968.

[24] Piene R., Polar Classes of Singular Varieties, *Ann. Sci. École Norm. Sup.* **11** (1978), pp. 247–276.

[25] Sabbah C., Morphismes analytiques stratifiés sans éclatement et cycles évanescents, *C. R. Acad. Sci. Paris Sér. I* **294** (1982), pp. 39–41, et article dans: *Analyse et topologie sur les espaces singuliers (II–III), 6–10 juillet 1981, Astérisque* **101–102** (1983), pp. 286–319.

[26] Teissier B., Cycles évanescents, sections planes et conditions de Whitney. In: *Singularités à Cargèse, 1972, Astérisque* **7-8** (1973), pp. 285–362.

[27] Teissier B., Variétés polaires I, *Invent. Math.* **40** (1977), pp. 267–292.

[28] Teissier B., Variétés polaires II. In: *Algebraic Geometry, Proceedings, La Rábida, 1981*, Lecture Notes in Math., 961, Springer, pp. 314–491.

[29] Teissier B., Résolution simultanée et cycles évanescents II, appendice. In: *Séminaire sur les singularités des surfaces*, Lecture Notes in Math., 777, Springer, pp. 136–143.

[30] Thom R., Ensembles et morphismes stratifiés, *Bull. Amer. Math. Soc.* **75** (1969), pp. 240–284.

[31] Verdier J.-L., Stratifications de Whitney et théorème de Bertini–Sard, *Invent. Math.* **36** (1976), pp. 295–312.

[32] Whitney H., Tangents to an Analytic Variety, *Ann. of Math.* **81** (1965), pp. 496–549.

[33] Zariski O., Some Open Questions in the Theory of Singularities, *Bull. Amer. Math. Soc.* **77** (1971), pp. 481–491.

Ajouté à la correction des épreuves:

[34] Hardt R., Stratification of Real Analytic Mappings and Images, *Invent. Math.* **28** (1975), pp. 193–208.

[35] Urabe T., Duality of Numerical Characters of Polar Loci, *Publ. R. I. M. S. Kyoto Univ.* **17** (1981), pp. 331–345.

[36] Denkowska Z. et Wachta K., La sous-analycité de l'application tangente, *Bull. Acad. Pol. Sci.* **30** (1982), pp. 329–331.

CENTRE DE MATHÉMATIQUES DE L'ÉCOLE POLYTECHNIQUE, PLATEAU DE PALAISEAU, 91128 PALAISEAU CÉDEX, FRANCE
LABORATOIRE ASSOCIÉ AU C.N.R.S. N° 169

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

W. BARTH

Report on Vector Bundles

In this paper I want to give a survey of the present state of the theory of algebraic vector bundles over compact base spaces. It is impossible to mention here all the recent publications on this subject. So I chose for discussing only those aspects which seem most interesting to me.

1. Stability

When a group acts on a variety the quotient variety exists for the set of stable points only. Invariant theorists of the last century were already familiar with this phenomenon, for which Mumford [16] gave a modern treatment. He recognized the importance of this concept for the construction of moduli spaces. (They usually occur as some quotient variety.) The following definition is essentially due to him: Let V be a bundle on a variety X and $H \subset X$ an ample divisor. Then V is *H-stable* (*in the sense of Mumford*) if for all subsheaves $S \subset V$ one has $\deg_H S/\operatorname{rank} S < \deg_H V/\operatorname{rank} V$, where $\deg_H S = c_1(S)\cdot H^{\dim X-1}$. This definition was applied to constructing moduli spaces for stable vector bundles, first on curves, then in full generality [14], [5]. A disadvantage is the dependence on the polarization H.

Bogomolov [3] introduced a version of stability which is independent of polarization: The rank-r bundle V is *stable* (*in the sense of Bogomolov*) if for every representation ϱ of $GL(r, C)$ with $\det\varrho = 1$, any section in V^ϱ with a zero vanishes identically. This property implies H-semistability for all H, and if $\dim X = 2$ also the inequality $(r-1)c_1^2 \leqslant 2rc_2$. For a surface X of general type Bogomolov proved that T_X is stable, hence $c_1^2 \leqslant 4c_2$ holds, a historically very important estimate for Miyaoka's proof of the famous $c_1^2 \leqslant 3c_2$ inequality.

There is also a differential geometric version of stability due to Kobayashi [12]. A hermitian metric h on V is called *hermite-einstein* if there is a hermitian metric g on X such that the curvature tensor θ of h satisfies $g_{ij}\theta_{\alpha\beta ji} = \varphi\delta_{\alpha\beta}$ with some real function $\varphi \geqslant 0$ or $\leqslant 0$ on X. The existence of a hermite-einstein metric on V over kähler X implies that V is not unstable in the sense of Mumford (with H being the kähler class). One feels that the existence of such a metric should characterize stability in a certain sense, but up to now one knows its existence only for certain stable bundles on curves, for $T_{\boldsymbol{P}_n}$ and for the null-correlation bundle on $\boldsymbol{P}_3$ [13].

When dealing with stable bundles on $\boldsymbol{P}_n$ one meets a variety of stability phenomena, for which a systematic group theoretic approach should be possible, but at the moment is still missing. Here are three examples:

(a) Stable bundles have a strong tendency to keep their stability under restriction to general subspaces. The most popular result of this sort is the *Grauert–Mülich theorem* [18]: If $V \mid \boldsymbol{P}_n$ is stable and $L \subset \boldsymbol{P}_n$ a general line, then $V \mid L = \bigoplus\limits_{1}^{r} \mathcal{O}_L(k_i)$, with integers $k_1 \leqslant k_2 \leqslant \ldots \leqslant k^r$ satisfying the *connectedness property*: $k_{i+1} \leqslant k_i + 1$.

(b) For V on $\boldsymbol{P}_3$ there are integers k_i determining to a large extent the dimensions $h^q(V(l))$, $l \in \boldsymbol{Z}$. This set of integers, called by Tjurin the *spectrum* of V, e.g. for stable V of rank 2 is again *connected*.

(c) Stable bundles on $\boldsymbol{P}_2$ with $c_1(V) = 0$ and $c_2(V) = k$ are characterized by a triplet a_0, a_1, a_2 of $k \times k$-matrices [10]. For rank $V \leqslant 3$ it is known (and in general it is expected) that these three matrices, under the natural $\mathrm{GL}(k, \boldsymbol{C}) \times \mathrm{GL}(k, \boldsymbol{C})$-action, cannot be simultaneously transformed into block form

$$\begin{bmatrix} 0 & s \times s \\ (k-s) \times (k-s) & * \end{bmatrix}, \quad 0 < s < k,$$

which means the triplet is stable under this group action.

2. Special base spaces

The description of bundles of rank $r \geqslant 2$ as elements of the cohomology set $H^1(X, \mathcal{Gl}(r, \boldsymbol{C}))$ is generally useless. Except for the concept of stability there is no general concept applicable to classifying bundles of arbitrary rank over arbitrary X. The classification depends on X very much and interesting results are known only in very few special cases:

(a) Compact Riemann surfaces. The systematic study of rank-2 bundles over one-dimensional X was started in 1955 by Atiyah. Bundles over curves are accessible because they are extensions of lower rank bundles by line bundles. In this way Atiyah classified rank-2 bundles over curves of genus $\leqslant 2$ and Tjurin over curves of an arbitrary genus. After Mumford provided stability, the moduli space $M(r, g, d)$ of stable rank-r bundles with determinant bundle of degree d over a curve of genus g was constructed and described by several authors [17]. The most beautiful and explicit result is the following one. Let X be a *hyperelliptic curve* of genus $g \geqslant 2$ and let $\lambda_1, \ldots, \lambda_{2g+2}$ be the affine coordinates of the branch points of a representation of X as a double cover of the projective line. It is the idea of A. Weil to relate the properties of X with those of the intersection of the two quadrics

$$Q_1\colon \sum_1^{2g+2} z_i^2 = 0, \quad Q_2\colon \sum_1^{2g+2} \lambda_i z_i^2 = 0$$

in $\boldsymbol{P}_{2g+1}$. In his 1973 thesis M. Reid proved that the jacobian of X is isomorphic to the variety of $(g-1)$-dimensional linear subspaces in $\boldsymbol{P}_{2g+1}$ contained in the intersection $Q_1 \cap Q_2$. In 1976 Desale and Ramanan [4] showed that the variety of $(g-2)$-dimensional linear subspaces contained in $Q_1 \cap Q_2$ can be identified with the *moduli space of stable rank-2 vector bundles* on X with a fixed determinant bundle of *odd* degree.

(b) The projective plane $\boldsymbol{P}_2$. Rank-2 bundles over surfaces are most conveniently described by means of the point sets which arise as zeros of general sections of the bundles. In this way Schwarzenberger (1961) constructed many examples, Maruyama proved the irreducibility of the moduli space of rank-2 bundles over $\boldsymbol{P}_2$, and Hoppe and Spindler (1980) and Brosius (1982) over ruled surfaces. This method meets its limits already on threefolds, because it uses information on subvarieties to study bundles instead of vice versa. The following method, going back to Horrocks, in my opinion is more promising: A bundle V on $\boldsymbol{P}_2$ is — up to direct summands of rank 1 — uniquely determined by $\bigoplus_k H^1\big(V(k)\big)$, considered as a graded module over the graded ring $\bigoplus_k H^0\big(\mathcal{O}_{\boldsymbol{P}_2}(k)\big)$ of homogeneous polynomials. If V is *stable* and $c_1(V) = 0$ already the "middle piece" of this module, namely the multiplication $H^0\big(\mathcal{O}_{\boldsymbol{P}_2}(1)\big) \otimes H^1\big(V(-2)\big) \to H^1\big(V(-1)\big)$ suffices. It is given by three $k \times k$-matrices a_0, a_1, a_2 with $k = c_2(V) = h^1\big(V(-2)\big) = h^1\big(V(-1)\big)$. The problem of this method (originating in its homological descendence) is that the rank r of V is given by the annoying formula

$$2k + r = \operatorname{rank}\begin{bmatrix} 0 & a_2 & -a_1 \\ -a_2 & 0 & a_0 \\ a_1 & -a_0 & 0 \end{bmatrix}.$$

But the description of all matrix triplets a_0, a_1, a_2 for fixed r is not too hard and can be used to prove the irreducibility and rationality of the moduli space of stable rank-r bundles over $\boldsymbol{P}_2$ with $c_1 = 0$ and $c_2 = k$ [10].

(c) Abelian surfaces. Here S. Mukai announced (in Japanese [15]) some special results and stated some conjectures which seem to open the road to an understanding of the classification.

3. The instanton connection

In 1977 instanton (or pseudoparticle) solutions of the euclidean Yang–Mills equation were identified with certain vector bundles over $\boldsymbol{P}_3$. It was shown [1] how to obtain the potentials and fields from matrices describing the bundles. However, up to now, nobody has succeeded in parametrizing the $(8k-3)$-dimensional manifold of $\mathrm{SU}(2)$-instantons (instanton number k) with $8k-3$ real parameters. The AHDM-matrices contain more parameters, subject to certain nonlinear polynomial constraints if they define an instanton (or vector bundle). This difficulty has its analogue in the description of the complexification of the instanton manifold, i.e., the moduli space of the so-called *mathematical instanton bundles* on $\boldsymbol{P}_3$. At the moment, for the spaces of such rank-2 bundles with $c_2 = k \geqslant 5$ one does not know if they are smooth or connected. Even the dimension is not known, and rationality is proved only for $k \leqslant 3$. It is conceivable (although not expected) that this space has components of complex dimension $> 8k-3$. The space is most conveniently described as $\mathrm{GL}(k, \boldsymbol{C}) \times \mathrm{Sp}(2k+2, \boldsymbol{C})$-quotient (under the natural action) of the space of *rigid instantons*. These are bundles V with an additional structure, e.g. an identification $H^1(V(-1)) = \boldsymbol{C}^k$ [20]. They are parametrized by eight AHDM-type complex $(k+1) \times k$-matrices A_i, B_i, $i = 0, 1, 2, 3$, satisfying (among other Zariski-open conditions) the following system of $5(k^2-k)$ nonlinear equations:

$$A_i^t B_j - B_i^t A_j + A_j^t B_i - B_j^t A_i = 0, \quad i, j = 0, 1, 2, 3.$$

A quick count of constants (subtracting from the number $8(k^2+k)$ of complex parameters the number $5(k^2-k)$ of constraints and $3k^2+5k+$ $+3 = \dim \mathrm{GL}(k, C) + \dim \mathrm{Sp}(2k+2, C)$ for the group action) yields the instanton dimension $8k-3$. This is a strong hint that the quadratic equations above are not dependent too much. On the other hand, however, if they were so independent as to leave us with a smooth intersection of $5(k^2-k)$ quadrics in $\boldsymbol{P}_{8(k^2+k)-1}$, already for $k > 9$ this space of rigid instantons would be a variety of general type, hence not unirational and hard to parametrize.

So at present, despite serious efforts, the algebraic vector bundle side of the instanton connection seems to be at a stop, unlike the use of Yang–Mills equations in other situations. Think of Donaldson's spectacular application of them to topology or of the increasing list of other solutions, e.g.

— merons,

— monopoles (translated into the language of analytic vector bundles by Hitchin [7]),

— periodic instantons (constructed by t'Hooft [8] on 4-dimensional real tori with quite suspiciously looking rational ratios between their periods).

4. Geometric aspects

The correspondence between line bundles over a variety and divisors in it is the most useful tool in algebraic geometry. As for bundles of higher rank, this correspondence is generalized in different ways:

(a) Sections in 2-bundles and subcanonical codimension-2 submanifolds. Serre was the first to analyze the zero-scheme $Z \subset X$ of a section s in a rank-2 bundle V on X. If $\operatorname{codim}_X Z = 2$, one has an exact sequence

$$0 \to \mathcal{O}_X \xrightarrow{s} V \to \mathcal{I}_Z \cdot \det V \to 0$$

and Z is subcanonical in the sense that ω_Z is the restriction to Z of some line bundle (namely $L = \omega_X \otimes \det V$) on X. Serre's technique inverting this construction (starting with a subcanonical Z and L to find V, provided the obstruction in $H^2(\omega_X \otimes L^*)$ vanishes) is nowadays standard. This correspondence has been most fruitful on $\boldsymbol{P}_3$, where it led e.g. to the following result.

A rank-2 bundle V on $\boldsymbol{P}_3$ has a section if $\chi(V) > 0$ and $c_2(V) > 0$ (Hartshorne [6]).

Due to lack of examples there was no comparable activity on $\boldsymbol{P}_n$, $n \geqslant 4$. The only accessible subcanonical surface in $\boldsymbol{P}_4$ (not a complete intersection) is still the torus of Horrocks–Mumford [9]. As it seems hard to describe its embedding independently of the existence of the corresponding bundle, there was a search for singular subcanonical surfaces in $\boldsymbol{P}_4$. There are some already described in [9]: Configurations of quadrics or the tangent scroll of an elliptic quintic curve. Another type [2] is a certain multiplicity-2 structure on a cone over a space curve $R \subset \boldsymbol{P}_3$, twisted along the generators of the cone. This structure is expected to exist for each 2-bundle on $\boldsymbol{P}_4$, but its construction depends on delicate properties of R. For the bundle [9] R is the curve (of degree 8) of contact of two Kummer surfaces having all their nodes on R, such that the nodal divisor D satisfies $\mathcal{O}_R(3D) = \mathcal{O}_R(6)$.

A construction for such an R not involving the bundle has not been found yet. The next simple case of such an R should be of degree 18 on a sextic surface with 48 even nodes. (Even means 2-divisibility of the divisor consisting of the blown up nodes on the nonsingular model of the surface. Although there is a large classical literature concerning this subject, such a particular sextic is unknown.)

Much of the recent interest in bundles over $\boldsymbol{P}_n$ comes from the *complete intersection problem* which, in its simplest form, is the question: Are there codimension-2 submanifolds in $\boldsymbol{P}_6$ other than complete intersections? The answer is still unknown.

(b) Morphisms of bundles and their degeneration sets. The classical subject of surfaces with many nodes and curves of contact has a vector bundle approach, too. Given V on $\boldsymbol{P}_3$ and a quadratic form on V (i.e., a selfadjoint map $q\colon V \to V^*$), the degeneration set $Y = \{\det q = 0\}$ is, in general, a surface with $4\big(c_1(V^*)c_2(V^*) - c_3(V^*)\big)$ even nodes where $\operatorname{rank} q = \operatorname{rank} V - 2$. Sections in V^* generate on Y curves of contact with other surfaces. The uniqueness question (how far are q and V determined by Y and S?) was investigated by Tjurin [19]; the existence question (given Y and S, do they come from some q?) is unsolved.

Now this concept can be widely varied: anti-symmetry of q or no symmetry at all, replace V^* by $V^* \otimes L$ for some line bundle L, or by a totally new vector bundle W. In this way e.g. Tjurin [20] associates with a rigid instanton, $c_2 = k$, a surface $D \subset \boldsymbol{P}_3$ of degree $2k$ with $4\binom{k+1}{3}$ nodes and two curves A, $B \subset D$ of degree $\binom{k+1}{2}$ such that the divisor $2(A-B)$ is equivalent (on the nonsingular model of D) to some linear combination of the blown up nodes.

(c) Normal bundles of space curves. Knowing rank-2 bundles on curves fairly well, it is natural to ask for the identification of the normal bundle N_C of a space curve $C \subset \boldsymbol{P}_3$ (which in the case of subcanonical C is the restriction to C of some 2-bundle on $\boldsymbol{P}_3$). In particular, Grauert asked to find some C with N_C not the direct sum of two line bundles. The first such example (a quintic genus-2 curve) was found by Van de Ven [21]. It stimulated recent work giving now quite complete information for rational and elliptic C. For a survey see [11].

Added in proof: Donaldson constructed hermite-einstein metrics on stable bundles over surfaces. He also solved the classification problem for instantons.

References

[1] Atiyah M. F., Hitchin N., Drinfeld V., Manin Ju., Construction of Instantons, *Phys. Lett.* **65 A** (1978), pp. 185–187.

[2] Barth W., Kummer Surfaces Associated with the Horrocks–Mumford Bundle. In: *Géométrie algébrique*, A. Beauville (ed.), Angers, 1979.

[3] Bogomolov F. A., Holomorphic Tensors and Vector Bundles on Projective Varieties, *Engl. transl. Math. USSR Izv.* **13** (1979), pp. 499–555.

[4] Desale U. V., Ramanan S., Classification of Vector Bundles of Rank 2 on Hyperelliptic Curves, *Inv. Math* **38** (1976), pp. 161–185.

[5] Forster O., Hirschowitz A., Schneider M., Type de scindage généralisé pour les fibrés stables. In: *Vector Bundles and Differential Equations*, Nice, 1979.

[6] Hartshorne R., Stable Reflexive Sheaves II, *Inv. Math.* **66** (1982), pp. 165–190.

[7] Hitchin N., Monopoles and Geodesics, *Comm. M. Phys.* **83** (1982), pp. 579–602.

[8] t'Hooft G., Some Twisted Self–Dual Solutions for the Yang–Mills Equations on a Hypertorus, *Comm. M. Phys.* **81** (1981), pp. 267–275.

[9] Horrocks G., Mumford D., A Vector Bundle on $\boldsymbol{P}^4$ with 15 000 Symmetries, *Topology* **12** (1973), pp. 63–81.

[10] Hulek K., On the Classification of Stable Rank-r Vector Bundles over the Projective Plane. In: *Vector Bundles and Differential Equations*, Nice, 1979.

[11] Hulek K., On the Normal Bundle of Space Curves. To appear in *Proc. Top. Sem. Mexico*.

[12] Kobayashi S., First Chern Class and Holomorphic Tensor Fields, *Nagoya M. J.* **77** (1980), pp. 5–11.

[13] Lübke M., Hermite–Einstein-Vektorbündel, Diss., Bayreuth, 1982.

[14] Maruyama M., Moduli of Stable Sheaves I and II, *J. M Kyoto Univ.* **17** (1977), pp. 91–126, and **18** (1978), pp. 557–614.

[15] Mukai S., On Classification of Vector Bundles over Abelian Surfaces, *Proc. Sym. Res. Inst. M. Sci.*, Kyoto Univ. 1980, *RIMS Kokyuroku* **409**, pp. 103–127.

[16] Mumford D., Geometric Invariant Theory, *Erg. d. M. n. F.* **34** (1965), Springer.

[17] Narasimhan M. S., Ramanan S., Moduli of Vector Bundles on a Compact Riemann Surface, *Ann. M.* **89** (1969), pp. 19–51.

[18] Spindler H., Der Satz von Grauert–Mülich ..., *Math. Ann.* **243** (1979), pp. 131–141.

[19] Tjurin A. N., The Geometry of Singularities of a Generic Quadratic Form, Engl. transl. in: *Math. USSR Izv.* **17** (1981).

[20] Tjurin A. N., *The Instanton Equations for* $(n+1)$-*Superpositions of Marked* ad $T\boldsymbol{P}^3 \oplus$ ad $T\boldsymbol{P}^3$, to appear.

[21] Van de Ven A., Le fibré normal d'une courbe dans $\boldsymbol{P}_3$ ne se décompose pas toujours, *C. R. Ac. Sci. Paris* **289 A** (1979), pp. 111–113.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

JOHN ERIK FORNÆSS

Holomorphic Mappings between Pseudoconvex Domains

Let Ω_1, Ω_2 be two bounded domains in C^n or in some complex manifold. A holomorphic map $F: \Omega_1 \to \Omega_2$ is said to be *proper* if $F^{-1}(K)$ is a compact subset of Ω_1 whenever K is a compact subset of Ω_2. A special case occurs when F is biholomorphic.

PROBLEM. Prove that F extends to a C^∞ map $\bar{F}: \bar{\Omega}_1 \to \bar{\Omega}_2$.

For references, see: K. Diederich, J. E. Fornæss, R. P. Pflug, *Convexity in Complex Analysis*, Springer-Verlag.

Some conditions on Ω_1, Ω_2 are necessary: Otherwise, let T be the torus $C/\mathscr{L}$ where $\mathscr{L}$ is some lattice in C. Let $\Omega_1 = \Delta \times T = \Omega_2 \subset C \times T$. (Here Δ is the unit disc in C.) Then there are lots of biholomorphic maps which are singular at the boundary. Let $\varphi(z)$ be an arbitrary holomorphic function on the unit disc. Then $(z, w) \to \big(z, w+\varphi(z)\big)$ is a biholomorphic map: $\Delta \times C \to \Delta \times C$ whose quotient under $\mathscr{L}$ becomes a biholomorphic map: $\Omega_1 \to \Omega_2$. Observe that Ω_1 is a pseudoconvex domain with real analytic boundary. Also observe that $\partial\Omega_1$ is Levi-flat.

We will now discuss the question in C^n.

PROBLEM. Let $\Omega_1, \Omega_2 \Subset C^n$ be bounded domains with C^∞ boundary, and $F: \Omega_1 \to \Omega_2$ a proper, holomorphic mapping. Does F extend to a C^∞ map $\bar{F}: \bar{\Omega}_1 \to \bar{\Omega}_2$?

There is no known counterexample to this question. For domains with large automorphism groups this problem was studied by H. Cartan, W. Kaup, Matsushima, Ochiai and others.

We will discuss the case where the automorphism group is not necessarily large.

The first result is from about 1973 when Henkin, Margulies and Vormoor proved Hölder continuity of order $1/2$ up to the boundary in the case when Ω_1, Ω_2 are strongly pseudoconvex. (Boundary continuity in the general case is still unknown.)

The ingredients in this proof are:

1. The boundary distance is preserved up to a constant multiple under F (Hopf lemma).

2. The Kobayashi-metric $c(p; X)$ at a point $p \in \Omega_j$ in a direction X satisfies the estimate

$$\frac{|X|}{\sqrt{d(p_1 \partial\Omega_j)}} \lesssim c(p; X) \lesssim \frac{|X|}{d(p, \partial\Omega_j)},$$

where $d(p, \partial\Omega_j)$ is Euclidean boundary distance. This is a simple consequence of the definition of strong pseudo-convexity.

3. The distance-decreasing property of the Kobayashi-metric under holomorphic mappings.

These facts give that

$$\frac{|X|}{d(p, \partial\Omega_1)} \gtrsim c_{\Omega_1}(p; X) \geqslant c_{\Omega_2}\big(f(p); f'(p)(X)\big) \gtrsim \frac{|f'(p)(X)|}{\sqrt{d(f(p), \partial\Omega_2)}}$$

hence $|f'(p)| \lesssim \dfrac{1}{\sqrt{d(p, \partial\Omega_1)}}$ from which Hölder 1/2 continuity easily follows by an integration argument.

The next major development was Fefferman's theorem in 1974.

THEOREM. Let Ω_1, Ω_2 *be strictly pseudoconvex domains with* C^∞ *boundary in* C^n. *Then if* F: $\Omega_1 \to \Omega_2$ *is biholomorphic,* F *extends to a* C^∞ *map* $\bar{F}$: $\bar{\Omega}_1 \to \bar{\Omega}_2$.

The ingredients in this proof are:

1. Careful analysis of the boundary behavior of the Bergman kernel using the fact that the Bergman kernel is explicitly known on the ball and the fact that strongly pseudoconvex domains can be approximated to high order by balls.

2. Using this to study the smoothness of the boundary behavior of geodesics of the Bergman metric and using the fact that the Bergman metric is invariant under biholomorphic mappings.

This proof has later been simplified. The first simplification was done by Sid Webster who replaced the second part of Fefferman's proof with a much simpler argument using the Bergman kernel on and off the diagonal instead of the Bergman metric. A similar simplification was also obtained by Ewa Ligocka, and later further simplifications were obtained by Steven Bell.

Bell introduced the important condition R: If $H_2(\Omega)$ denotes the space of L^2-functions on Ω which are holomorphic on Ω, then H^2 is a Hilbert subspace of L^2.

DEFINITION. Ω satisfies condition R if the orthogonal projection P_Ω: $L^2 \to H_2$ satisfies the property that $P_\Omega(C^\infty(\bar{\Omega})) \subset C^\infty(\bar{\Omega}) \cap H_2$.

There are not known any bounded domains with C^∞ boundary in C^n for which condition R fails.* Condition R is known to hold for pseudoconvex domains of finite type. This includes strongly pseudoconvex domains and pseudoconvex domains with real analytic boundary. Condition R is also known to hold for certain nonpseudoconvex domains with sufficiently many symmetries, for example complete Reinhardt domains (Bell, Boas, Barrett). Here the idea is that on such domains the Bergman kernel $K(z, w)$ is a function of the product $z\bar{w}$ and therefore, if w is restricted to a compact subset, K is holomorphic in z in a strictly larger domain.

THEOREM (Bell). *Let Ω_1, $\Omega_2 \Subset C^n$ be pseudoconvex domains with C^∞ boundary and assume Ω_1 satisfies condition* R. *If F: $\Omega_1 \to \Omega_2$ is a proper holomorphic mapping then $u\cdot(h\circ F) \in A^\infty(\Omega_1)$ whenever $h \in A^\infty(\Omega_2)$.*

Here $A^\infty(\Omega_j)$ are the C^∞ functions on $\bar{\Omega}_j$ which are holomorphic on Ω_j. Also u is the determinant of the complex Jacobian of F. The two most important special cases are

(i) $h \equiv 1$. This gives $u \in A^\infty(\Omega_1)$.

(ii) $h = w_j$, the jth coordinate function on the target space. Then $F_j \cdot u \in A^\infty(\Omega_1)$.

Of course, if the F_j's are in $A^\infty(\Omega_1)$, then we have the C^∞ extendibility of F.

To prove this one needs

1. The boundary distances are preserved up to finite order: $\exists \varepsilon > 0$ such that

$$d(p, \partial\Omega_1) \lesssim d^\varepsilon(f(p), \partial\Omega_2), \quad d(f(p), \partial\Omega_2) \lesssim d^\varepsilon(p, \partial\Omega_1).$$

This follows from the fact that pseudoconvex domains have Hölder continuous exhaustion functions and the Hopf lemma.

2. The transformation rule for the Berman projection:

$$P_1(u\cdot h\circ F) = u\cdot((P_2 h)\circ F).$$

* *Note added in proof*: Barret, Irregularity of the Bergman Projection on a Smooth Domain in C^2, *Ann. of Math.* **119** (1984), pp. 431–436, has found a (nonpseudoconvex) counterexample.

This is immediate in the case F is biholomorphic because $H_2(\Omega_1)$ and $H_2(\Omega_2)$ are isometric via $h \in H_2(\Omega_2)$ corresponding to $u(h\circ F) \in H_2(\Omega_1)$. For the case when F is a proper map this was a rather surprising property discovered by Bell. Observe that in this case however, the formula is immediate at least for holomorphic functions.

3. Another discovery by Bell is that if $h \in A^\infty(\Omega_2)$ and k is any positive integer, then there exists an $\tilde{h} \in C^\infty(\bar{\Omega}_2)$ such that $P_2\tilde{h} = h$ and $\tilde{h}$ vanishes to kth order at $\partial\Omega_2$.

The key observation used inductively in this proof is that if $g \in C^\infty(\bar{\Omega}_2)$ and $g \mid \partial\Omega_2 = 0$, then $\partial g/\partial z_j \perp H_2(\Omega_2)$. If $f \in H_2(\Omega_2)$, then $\langle f, \partial g/\partial z_j\rangle = \int_{\Omega_2} f\overline{\partial g/\partial z_j} = \int \partial f/\partial \bar{z}_j \cdot \bar{g} = \int 0 = 0$. Hence

$$h - \partial g/\partial z_j \in C^\infty(\bar{\Omega}_2), \qquad P_2(h - \partial g/\partial z_j) = P_2 h = h.$$

4. The Jacobian u blows up to at most finite order at $\partial\Omega_1$. This is just Cauchy's integral formula.

Applying these facts: let $h \in A^\infty(\Omega_2)$. Then choose k large and let $\tilde{h} \in C^\infty(\bar{\Omega}_2)$ so that $\tilde{h}$ vanishes on $\partial\Omega_2$ to order k and $P_2\tilde{h} = h$. This implies that for some large $k' \leqslant k$ the function $u\cdot\tilde{h}\circ F$ is in $C^{k'}(\bar{\Omega}_1)$. Condition R implies (by the closed graph theorem) that for some large $k'' \leqslant k'$, $P_1(u\cdot\tilde{h}\circ F) \in C^{k''}(\bar{\Omega}_1)$. Here $k'' \to \infty$ when $k \to \infty$. But observe that $P_1\big(u\cdot(\tilde{h}\circ F)\big) = u\cdot(P_2\tilde{h}\circ F) = u(h\circ F)$. Hence $u(h\circ F) \in A^\infty(\bar{\Omega}_1)$.

THEOREM (Bell–Catlin, Diederich–Fornæss). *Let Ω_1, Ω_2 be bounded pseudoconvex domains in C^n with C^∞ boundary and assume that Ω_1 satisfies condition* R. *If $f\colon \Omega_1 \to \Omega_2$ is a proper, holomorphic map then f extends to a C^∞ map $f\colon \bar{\Omega}_1 \to \bar{\Omega}_2$.*

The proof of this theorem is in two parts. At first one proves that the Jacobian u vanishes to finite order. And the second part is a division theorem that allows one to divide $u\cdot h\circ F$ by u to get $h\circ F \in A^\infty(\bar{\Omega}_1)$.

Bell has also shown that if $\Omega_1, \Omega_2 \Subset C^n$ are bounded domains with C^∞ boundary (without the assumption of pseudoconvexity) both of which satisfy condition R, then proper holomorphic maps extend smoothly up to the boundary. In this case it is also possible to show that $h \in A^\infty(\bar{\Omega}_2)$ implies $u\cdot h\circ F \in A^\infty(\Omega_1)$.

Furthermore, Bedford–Bell–Catlin has extended the theorem to the case of pseudoconvex domains in Stein manifolds with Ω_1 satisfying condition R. In that case the appropriate version of H_2 are the L^2-integrable $(n, 0)$-forms with holomorphic coefficients and these transform appropriately with respect to holomorphic mappings. The important point is that the inner product is independent of the choice of metric on the manifold.

Now we will discuss mappings defined in more general sets than subsets of C^n. Let M be a complex manifold of dimension $n+q$ for some $n \geqslant 3$ and $q \geqslant 0$. Let $\Omega \Subset M$ be a pseudoconvex domain with C^∞ boundary. We assume moreover that $\Omega = \{r < 0\}$ where r is a smooth plurisubharmonic defining function defined in a neighborhood of $\bar{\Omega}$. Also assume that the Levi-form of $\partial\Omega$ has at least $n-1$ strictly positive eigenvalues. Finally, let Ω' be a bounded strongly pseudoconvex domain with C^∞ boundary in C^n.

THEOREM (Diederich–Fornæss). *If F: $\Omega \to \Omega'$ is a proper holomorphic map, then F extends to a C^∞ map: $\bar{\Omega} \to \bar{\Omega}'$. Moreover, F has constant rank n near $\partial\Omega$.*

The prime example of such a situation occurs when Ω is a holomorphic fiber bundle with base Ω' and compact fiber. Observe that if F: $\Omega \to \Omega'$ is such a map, then we can blow up an arbitrary point $p \in \Omega$ to get another domain $\tilde{\Omega}$ with a blow-down map π: $\tilde{\Omega} \to \Omega$. Then $F \circ \pi$: $\tilde{\Omega} \to \Omega'$ is another proper holomorphic mapping. Of course, $F \circ \pi$ has rank 0 at $\pi^{-1}(p)$.

The main steps in the proof are the following:

1. One shows that Ω is foliated near $\partial\Omega$ by complex manifolds of dimension q. The leaves are connected components of level sets of F. This follows by using the special hypotheses on Ω. The fact that Ω has at least n positive eigenvalues makes it possible locally to take n-dimensional strongly pseudoconvex cross-sections. This implies that the level sets have dimension n. The level sets of r must also be pseudoconvex with exactly n strictly positive eigenvalues. This means that the level sets of r are foliated by n-dimensional complex manifolds. One then shows that the leaves of this foliation are connected components of the level sets of F. Notice that $\partial\Omega$ is also foliated by q-dimensional complex manifolds. We don't yet however know that these are compact.

2. The foliation has the following regularity property up to the boundary: The leaves have uniformly bounded volume and uniformly bounded diameter. For any leaf in the boundary, the holonomy group is finite and consists of biholomorphic maps of cross-sections, C^∞ up to the boundary. This uses that the leaves are essentially level sets of the holomorphic map F. Reason: The holonomy maps of the foliation at interior leaves can be controlled (uniformly Lipschitz). This is because the level sets of r are preserved under these holonomy maps. This forces control which allows one to estimate how the volumes and diameters of leaves change.

Once one has uniform control in the interior, this carries over to the boundary to give uniform bounds on volumes and diameters up to the boundary. The finiteness of the holonomy groups follows by application of the identity theorem for holomorphic mappings.

3. One next introduces the spaces of (n, p)-forms which are invariant under the holonomy map: Fix a leaf of the foliation. Then there exists a tubular neighborhood basis of the leaf each of which is closed under taking leaves. Moreover we may choose a cross-section around this leaf on which each element of the holonomy group acts as a biholomorphism. Starting with a smooth form on the cross-section and composing with all elements of the holonomy group one gets a form which is invariant under the taking of group-elements. Then this form can be pulled back to a form on the whole tubular neighborhood. Using partitions of unity one gets global spaces of such forms. Further, Sobolev norms can be introduced in these spaces, and also the $\bar{\partial}$- and the $\bar{\partial}^*$-operator. Hence the whole theory of the $\bar{\partial}$-Neumann operator can be carried over to this setting. And then one can apply the methods of the above theorems by Bell and others to complete the proof.

PRINCETON UNIVERSITY

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

REESE HARVEY

Calibrated Geometries*

1. Introduction

Our starting point is a classical demonstration that the shortest distance from one point to another in Euclidean space is the line segment M joining the two points. Let φ denote the constant coefficient differential form dual to the unit vector in the direction of the line segment M. Suppose M' is any other oriented curve joining the two points. Then

$$\operatorname{length}(M) = \int_M \varphi = \int_{M'} \varphi \leqslant \operatorname{length}(M'). \tag{1}$$

This simple idea provides the basic method for showing that a given submanifold M of a Riemannian manifold X is volume minimizing. The method can be axiomatized as follows. Suppose φ is a differential p-form on X with the following properties:

$$d\varphi = 0, \tag{2}$$

i.e., φ is closed under exterior differentiation, and

$$\operatorname{comass} \varphi \leqslant 1, \tag{3}$$

i.e., $\varphi|_W \leqslant \operatorname{vol}_W$ for all oriented p-dimensional subspaces W of the tangent space to X at all points of X. We call such a form φ a *calibration* and the pair (X, φ) is called a *calibrated manifold.* A calibration φ determines a geometry of submanifolds. Any oriented p-dimensional submanifold M with

$$\varphi|_M = \operatorname{vol}_M \tag{4}$$

will be called a *φ-submanifold.* The appropriate concept of M providing shortest distance or least volume takes the following form. Suppose,

* Written for the Proceedings of the International Congress of Mathematicians, Warsaw, August 1983.

for any other oriented p-dimensional submanifold M' with $M-M'$ homologous to zero (in particular, M' has the same boundary as M), that the volume of M' remains at least as large as the volume of M. Then M is said to be ***homologically volume minimizing.***

Now suppose that M is a φ-submanifold and that M' is a competitor for least volume with $M-M'$ homologous to zero. Then,

$$\mathrm{vol}(M) = \int_M \varphi = \int_{M'} \varphi \leqslant \int_M \mathrm{vol}(M'). \tag{1'}$$

The first equality is a consequence of (4), the second equality is a consequence of (2) and the homology of M' with M, while the inequality is a consequence of (3).

Thus, a calibration φ determines a geometry of submanifolds and each φ-submanifold is homologically volume minimizing.

It is convenient to describe the φ-submanifolds in the following framework. The metric allows us to identify an oriented p-plane (let $e_1, \ldots, e_p$ denote an orthonormal basis) with the simple/decomposable unit p-vector $\xi = e_1 \wedge \ldots \wedge e_p$ in $\Lambda^p T_x X$. Thus $G(p, T_x X)$, the grassmannian of oriented p-planes, is a natural subset of the space $\Lambda^p T_x X$ of p-vectors at $x \in X$. The calibration φ defines a hyperplane

$$\varphi = 1 \quad \text{in } \Lambda^p T_x X$$

and the grassmannian lies in the half space $\varphi \leqslant 1$ (this is condition (3) above). The contact set where the hyperplane $\varphi = 1$ intersects the grassmannian will be denoted $G_x(\varphi)$. Thus M is a φ-submanifold if and only if its unit oriented tangent space $\vec{M}_x$ belongs to the distinguished subset $G_x(\varphi)$ of the grassmannian at each point x of M.

There are many examples of calibrations and their associated geometries. Some are classical and some recent; some contain just one important φ-submanifold while others contain a rich variety of φ-submanifolds.

2. Hypersurface graphs

Suppose the ambient Riemannian manifold X is Euclidean space $\boldsymbol{R}^{n+1}$ and M is the graph of a scalar function f over a domain $\bar{D}$ with smooth boundary in $\boldsymbol{R}^n$. Then, along M, there is only one choice for a form φ of comass $\leqslant 1$ with $\varphi|_M = \mathrm{vol}_M$. Namely,

$$\varphi \equiv (1+|\nabla f|^2)^{-1/2}\left(dx_1 \wedge \ldots \wedge dx_n + \sum_{j=1}^{n} (-1)^{n-1} \frac{\partial f}{\partial x_j} * dx_j\right).$$

Taking this as definition of φ on all of $\bar{D}\times\boldsymbol{R}$, the condition, $d\varphi = 0$, is just the minimal surface equation

$$\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_j}(1+|\nabla f|^2)^{-1/2}\right) = 0.$$

This equation insures that φ is a calibration and hence that the φ-submanifold M is homologically volume minimizing in $\bar{D}\times\boldsymbol{R}$; in fact, ***absolutely*** volume minimizing since $\bar{D}\times\boldsymbol{R}$ has no homology in dimension n. This example is a special case of a calibrated foliation.

3. Calibrated foliations

Suppose X is a Riemannian manifold and φ is a calibration on X. If there is exactly one oriented subspace F_x of the tangent space at each point $x\in X$ such that $\varphi|_F = \mathrm{vol}_F$, and if this distribution F is integrable then the induced foliation $\mathscr{F}$ is called a ***calibrated foliation***. M is a φ-submanifold if and only if M is the union of pieces of leaves. Thus every closed leaf in a calibrated foliation is homologically area minimizing.

4. Kähler geometry (Calibrated complex manifolds)

Suppose X is a Kähler manifold with Kähler form ω and the induced hermitian metric. Take φ to be the closed form $\frac{1}{p!}\,\omega^p$. The classical Wirtinger inequality has two implications. First, φ has comass $\leqslant 1$ and hence φ is a calibration. Second, M is a φ-submanifold if and only if M is a p-dimensional complex submanifold. Thus every submanifold (more generally every complex subvariety) of a Kähler manifold is homologically volume minimizing. This is the classical result of Federer [3]. We will return to Kähler geometry from this point of view later.

5. Special Lagrangian geometry

This example provides a new geometry to study (see Harvey–Lawson [10]). Suppose X is a hermitian manifold with trivial canonical bundle. Let Ω denote a never vanishing holomorphic n-form. Assume that $|\Omega|\equiv 1$ (change the metric by a conformal factor if necessary). Then

$$\varphi\equiv \mathrm{Re}\,\Omega$$

is a calibration on X.

We will discuss, in further detail, the case where $X \equiv C^n = R^n \oplus iR^n$ is complex Euclidean space with the standard Kähler metric and

$$\varphi \equiv \operatorname{Re} dz_1 \wedge \ldots \wedge dz_n.$$

There are many analogues to the developments in the Kähler case.

First, there is an inequality analogous to the Wirtinger inequality. Recall that the Lagrangian n-planes, which are just the n-planes on which the Kähler/symplectic form vanishes, can be described as the images of $R^n \subset C^n$ under the action of the unitary group U_n. *The special Lagrangian planes* are just the images of $R^n \subset C^n$ under the action of the special unitary group SU_n. The special Lagrangian inequality says:

$$(\operatorname{Re} dz)(\xi) \leqslant 1 \quad \text{for all } \xi \in G(n, C^n),$$

with equality if and only if ξ is special Lagrangian.

Second, there is an analogue to the Cauchy–Riemann equations. Suppose $M \subset C^n$ is the graph of a function F from R^n to R^n. Recall that M is Lagrangian if and only if $F \equiv \nabla f$ for some real-valued potential or generating function f. The graph M is special Lagrangian if and only if the alternating sum of the odd elementary symmetric functions of the hessian of f vanishes. This special Lagrangian differential equation is simply

$$\Delta f = \det(\operatorname{Hess} f)$$

when $n = 3$.

Third, in analogy with the geometry of complex submanifolds in C^n, there is a rich variety of examples, both of a specific nature and as general classes. We mention just one example. Suppose n is odd. Define two disjoint $(n-1)$-dimensional tori T^+ and T^- by

$$T^+ \equiv \{(e^{i\theta_1}, \ldots, e^{i\theta_n}) \colon \theta_1 + \ldots + \theta_n = 0\},$$

$$T^- \equiv \{(e^{i\theta_1}, \ldots, e^{i\theta_n}) \colon \theta_1 + \ldots + \theta_n = \pi\}.$$

Let $M^\pm$ denote the cone with vertex at the origin through $T^\pm$. Note that M^- is the antipodal image of M^+. Thus M^+ cannot be a real analytic variety. However M^+ is special Lagrangian and hence absolutely area minimizing. This is perhaps the simplest example for an absolutely area minimizing cone in Euclidean space which is not a real-analytic variety.

A harmonic gradient map $F = \nabla f$ with degenerate image has $\det(\operatorname{Hess} f) = 0$. Thus, in dimension 3, the study of harmonic gradient maps is a special case of special Lagrangian geometry. This clarifies the appearance of the minimal surface equation in the beautiful work of Hans Lewy [18] on harmonic gradient maps. See Harvey and Lawson [10] for more details on this class of "degenerate" special Lagrangian 3-folds in $\boldsymbol{R}^6$ as well as for many other examples.

6. Exceptional geometries

One of our earliest discoveries (Harvey and Lawson [10]) in this investigation was the appearance of certain beautiful "exceptional" geometries in low dimensions. These geometries are associated with the octonians or Cayley numbers $\boldsymbol{O} \cong \boldsymbol{R}^8$. Although Cayley multiplication is not associative, there is a weak form of associativity. Namely, the associator

$$[a, b, c] \equiv (ab)c - a(bc)$$

vanishes if any two of the octonians a, b, c are equal. Thus the associator is alternating. Consequently, we have a well-defined notion of ***associativity*** for a 3-plane in, say, the purely imaginary octonians $\mathrm{Im}\,\boldsymbol{O}$ ($\mathrm{Im}\,\boldsymbol{O}$ is the hyperplane orthogonal to the multiplicative unit $1 \in \boldsymbol{O}$). Namely, the associator should vanish for each (or equivalently all) basis a, b, c for the 3-plane. The associative 3-planes can be characterized as the imaginary parts of quaternion subalgebras of $\boldsymbol{O}$.

Now we wish to describe the geometry based on the associative 3-form φ on $\mathrm{Im}\,\boldsymbol{O}$ defined by:

$$\varphi(a, b, c) = \langle a, bc \rangle$$

using the standard inner product $\langle\ ,\ \rangle$ on $\boldsymbol{O}$. As with special Lagrangian geometry we outline the analogues to first, the Wirtinger inequality, second, the Cauchy–Riemann equation, and third, a rich class of examples.

The *Associator inequality* says that:

$$\varphi(\xi) \leqslant 1 \quad \text{for all } \xi \in G(3, \mathrm{Im}\,\boldsymbol{O}),$$

and furthermore,

$$\varphi(\xi) = 1 \Leftrightarrow \xi \text{ is associative and canonically oriented by its quaternionic structure.}$$

Thus φ is a calibration on $\boldsymbol{R}^7 \cong \mathrm{Im}\,\boldsymbol{O}$. The φ-grassmannian $G(\varphi)$, which consists of all associative 3-planes, can be naturally represented as the quotient G_2/SO_4.

It is illuminating to complete the associator inequality to an equality:

$$\langle a, bc\rangle^2 + \tfrac{1}{4}\|[a, b, c]\|^2 = \|a\wedge b\wedge c\|^2 \quad \text{for all } a, b, c \in \operatorname{Im}\boldsymbol{O}.$$

Now the differential equations governing this geometry can be deduced from the vanishing of the associator.

Let $\boldsymbol{H}$ denote a (standard) quaternion subalgebra of $\boldsymbol{O}$ and make the identification $\operatorname{Im}\boldsymbol{O} = \operatorname{Im}\boldsymbol{H}\oplus\boldsymbol{H}$. Suppose $M \subset \operatorname{Im}\boldsymbol{O}$ is the graph of a function from $\operatorname{Im}\boldsymbol{H}$ to $\boldsymbol{H}$ and let x_2, x_3, x_4 denote i, j, k coordinates on $\operatorname{Im}\boldsymbol{H}$ with respect to the standard i, j, k axes. The *Dirac operator* is defined by

$$D(f) = -\frac{\partial f}{\partial x_2}i - \frac{\partial f}{\partial x_3}j - \frac{\partial f}{\partial x_4}k;$$

and the *first order Monge–Ampère operator* is defined by

$$\sigma(f) \equiv \frac{\partial f}{\partial x_2}\times\frac{\partial f}{\partial x_3}\times\frac{\partial f}{\partial x_4},$$

with the aid of certain triple cross product defined by $a\times b\times c \equiv \frac{1}{2}\big(a(bc) - c(ba)\big)$.

The graph M is associative if and only if f satisfies the differential equation

$$D(f) = \sigma(f).$$

Third, there is a rich variety of examples of associative submanifolds. Perhaps the most interesting to mention are the associative cones. Their singular structure reflects the singular structure of general associative varieties. Suppose M is a 3-dimensional cone in $\operatorname{Im}\boldsymbol{O}$ with vertex at the origin. Let $\Gamma \equiv M\cap S^6$ denote the link in the unit sphere S^6 about the origin. There is a natural almost complex structure on $S^6 \subset \operatorname{Im}\boldsymbol{O}$. Simply define $J_x\colon T_xS^6\to T_xS^6$ by $J_xu \equiv xu$ (octonian multiplication). Now the cone M is associative if and only if the link Γ is a complex curve in S^6. Robert Bryant [1], in an exciting development, has shown that every Riemann surface appears as a complex curve in $S^6 \subset \operatorname{Im}\boldsymbol{O}$; and hence as the link on an associative cone.

There is a dual geometry of 4-folds in $\operatorname{Im}\boldsymbol{O}$, defined by the property that the normal 3-plane be associative, referred to as coassociative geometry. Here we are content to briefly mention that the famous Lawson–Osserman [17] example, of a Lipschitz solution to the non-parametric minimal surface equation which is not continuously differentiable, can be embedded in a family of coassociative submanifolds. In particular, this establishes that their example is absolutely area minimizing and not just minimal or stationary.

Finally, there is an exceptional geometry based on a Spin_7 invariant 4-form Φ on $\boldsymbol{O} \cong \boldsymbol{R}^8$ which includes all at once:

1. coassociative submanifolds in $\mathrm{Im}\,\boldsymbol{O} \subset \boldsymbol{O}$,
2. product manifolds $\boldsymbol{R} \times M$ where M is associative,
3. complex surfaces (negatively oriented) in $\boldsymbol{C}^4 \cong \boldsymbol{O}$ for a 6-sphere of complex structures,
4. special Lagrangian submanifolds of $\boldsymbol{C}^4 = \boldsymbol{R}^4 \oplus i\boldsymbol{R}^4 \cong \boldsymbol{O} = \boldsymbol{H} \oplus \boldsymbol{H}e$.

Again, analogues of (1) the Wirtinger inequality, (2) the Cauchy–Riemann equations, and (3) the rich variety of complex submanifolds of $\boldsymbol{C}^n$ are possible (see Harvey and Lawson [10]).

7. Double point geometry

An interesting question raised by Frank Morgan [21] is: When is the union of a pair of k-planes through the origin in Euclidean space absolutely area minimizing?

To prove a union of planes is absolutely area minimizing it suffices to find a calibration for the union. The simplest case, which does not follow from classical facts about $1, 2, n-1$, or $n-2$ forms, is the case of 3-forms in $\boldsymbol{R}^6$.

A pair ξ, η of three planes in $\boldsymbol{R}^6$ can be put in the following semi-canonical form with respect to an orthonormal basis $e_1, \ldots, e_6$:

$$\xi = e_1 \wedge e_2 \wedge e_3,$$

and

$$\eta \equiv (\cos\theta_1 e_1 + \sin\theta_1 e_4) \wedge (\cos\theta_2 e_2 + \sin\theta_2 e_5) \wedge (\cos\theta_3 e_3 + \sin\theta_3 e_6).$$

If either $\pm\theta = \pm(\theta_1, \theta_2, \theta_3)$ belongs to the tetrahedron with vertices $(0, 0, 0)$, $(0, \pi, \pi)$, $(\pi, 0, \pi)$, $(\pi, \pi, 0)$ then the pair $\{\xi, \eta\}$ is called a *double point*.

Morgan [21] has shown that each double point $\{\xi, \eta\} \subset G(3, \boldsymbol{R}^6)$ can be calibrated. That is, $\{\xi, \eta\} \subset G(\varphi)$ for some calibration $\varphi \in \Lambda^3(\boldsymbol{R}^6)^*$. Thus the union of the pair of planes for a double point pair is absolutely area minimizing in $\boldsymbol{R}^6$.

8. Coflat calibrations and Lawson–Simons cones

Federer [4] developed a beautiful and natural theory of coflat forms which enable one to utilize the basic idea, (1) above, when the calibration is only defined and smooth outside a closed set of Hausdorff p-measure zero. (Also in Harvey and Lawson [10], Theorem 4.9 there is a slight modification using a removable singularity theorem of Harvey and Polking [14]).

Earlier in Lawson [16] a number of interesting coflat calibrations of codegree 1 were essentially constructed by means of certain compact Lie group representations. Each of the resulting geometries of submanifolds contains a (single) cone of codimension one. Among these there is the Simons cone on $S^3 \times S^3 \subset \boldsymbol{R}^8$ which was originally proved to be absolutely area minimizing by Bombieri, De Giorgi, and Giusti.

9. Parallel calibrations on $\boldsymbol{R}^n$

An obvious and appealing class of calibrated geometries is provided by constant p-forms of comass one on Euclidean space $\boldsymbol{R}^n$. If $p = 1$ or $n-1$ then $G(\varphi)$ must be a singleton and hence the associated geometry is that of parallel lines or hyperplanes. If $p = 2$ or $p = n-2$ then $G(\varphi)$ is a complex grassmannian and we are in case 4 above.

The situation rapidly becomes more complicated. If $p = 3$ and $n = 6$ then a complete characterization of the possible geometries is given in Dadok and Harvey [2]. The geometry $G(\varphi)$ determined by φ is either

(1) special Lagrangian,

(2) complex crossed with a line,

(3) double point,

or

(4) singleton (i.e. a family of parallel 3-planes).

In particular, a pair $\{\xi, \eta\} \subset G(3, 6)$ can be calibrated if and only if $\pm\theta$ (the canonical angles) belongs to the tetrahedron described in Section 7; while a finite collection of three or more planes $\{\xi_1, \ldots, \xi_N\} \subset G(3, 6)$ can be calibrated if and only if $\xi_1, \ldots, \xi_N$ are all special Lagrangian for some special Lagrangian structure on $\boldsymbol{R}^6$.

If $p = 3$ and $n = 7$ then a complete characterization is again possible (see Harvey and Morgan [13]). In this case contact sets $G(\varphi)$ of the hyperplane $\varphi = 1$ with the grassmannian $G(3, 7)$ need not even be totally geodesic.

10. Calibrating foliations and complex manifolds

Suppose X is a compact smooth manifold equipped with an oriented p-dimensional smooth foliation. It is natural to ask

When can the foliation be calibrated?

That is, when does there exist a Riemannian metric on X and a calibration φ on X so that for each point x, $\varphi_x(\xi) = 1$ if (and only if) ξ is the unit oriented tangent space to the leaf through x?

There is an obvious purely topological necessary condition, since the calibration can be used to compute the volume of leaves. Each compact leaf, in fact every positive integral chain of compact leaves, must represent a non-zero homology class. If the foliation $\mathscr{F}$ is of codimension one then this necessary condition is also sufficient (Harvey and Lawson [9]). In higher codimension the following is necessary and sufficient to calibrate $\mathscr{F}$: Each non-zero foliation cycle represents a non-zero homology class (Harvey and Lawson [9]). A foliation cycle is a d-closed current T of finite mass such that $\vec{T}_x$ (the generalized tangent p-vector to T at x) coincides with $\vec{F}_x$ (the unit oriented tangent p-vector to the leaf at x) at almost all points x with respect to the generalized volume measure $\|T\|$ for T.

These results are closely related to work of D. Sullivan [13]. In particular, the Hahn–Banach step in the proof is due to Sullivan.

Suppose X is a compact complex manifold. It is natural to ask

When can the complex manifold be calibrated?

Interestingly, *this question is the same as asking*

When does a complex manifold admit a Kähler metric?

Again there is a natural necessary condition, since the Kähler form can be used to compute the volume of complex curves. Each compact complex curve, in fact each one-dimensional positive analytic cycle T, must represent a non-zero homology class. There are two ways this condition must be strengthened to become sufficient (Harvey and Lawson [12]). First, T must be allowed to be any positive d-closed current of complex dimension 1. Second, the conclusion that T represents a non-zero homology class must be replaced by:

T is not the bidimension $1, 1$ component of a boundary.

This condition is obviously necessary. Note that if $T = (dS)_{1,1}$ is the positive $1, 1$ component of a boundary and ω is a Kähler form then the mass of T equals $T(\omega) = (dS)_{1,1}(\omega) = (dS)(\omega) = S(d\omega) = 0$, so T must vanish. This condition is also sufficient (Harvey and Lawson [12]).

We will mention just two applications. Suppose X is an elliptic surface. Then the following are equivalent:

(1) X admits a Kähler metric,
(2) The first Betti number of X is even,
(3) The general fiber of X does not bound.

The equivalence of (1) and (2) is a result of Miyaoka [20].

Recently, an important application of Yau's solution to Calabi's conjecture (see Todorov [25] and Siu [22]) establishes that each K3 surface admits a Kähler metric. This K3 result and Miyaoka's theorem suffice, under Kodaira's classification, to establish that (1) and (2) are equivalent for all compact surfaces X.

Using methods that take no special note of K3 surfaces, it is shown in Harvey and Lawson [12] that each compact complex surface X with even first Betti number admits a "weakly" Kähler form ω. More precisely ω is a d-closed 2-form with positive definite 1, 1 component whose 2, 0 component (although not necessarily zero as in the Kähler case) is $\partial\alpha$ for some 1, 0 form α.

It is useful to interpret the Characterization Theorem of Harvey and Lawson [12] as a characterization of non-Kähler manifolds. It then asserts that each non-Kähler manifold carries an analytic object T, namely a positive bidimension 1, 1 current which is the 1, 1 component of a boundary. These objects appear to be intimately related to the geometry of the non-Kähler manifold (see Harvey and Lawson [12] for details).

For complex dimension 3 or more an interesting class of complex manifolds, strictly larger than Kähler, has been characterized by M. L. Michelsohn [19].

11. Boundaries

Perhaps the deepest open questions concerned with the geometries introduced in Harvey and Lawson [10] are about boundaries. The questions are analogous to a question which has been answered in the complex case (Harvey and Lawson [7]). The problem is this:

Characterize the boundaries of φ-submanifolds

(in the four cases, special Lagrangian, associative, coassociative, and Cayley).

The local real-analytic problem follows from the Cartan–Kähler theorem (see Harvey and Lawson [10] and [11]). A manifold is the boundary of a φ-submanifold if and only if it is maximally φ-like. This parallels the complex case. In addition, there are natural global restrictions on a candidate for a boundary (moment conditions).

Are these local and global necessary conditions also sufficient to ensure that a given submanifold is a φ-submanifold?

References

[1] Bryant R. L., Submanifolds and Special Structures on the Octonions, *J. Differential Geom.* **17** (1982).

[2] Dadok J., Harvey R., Calibrations on $\boldsymbol{R}^6$, *Duke Math. J.* **50** (1983), pp. 1231–1243.

[3] Federer H., Some Theorems on Integral Currents, *Trans. A.M.S.* **117** (1965), pp. 43–67.

[4] Federer H., Real Flat Chains, Cochains and Variational Problems, *Indiana Univ. Math. J.* **24** (1974), pp. 351–407.

[5] Federer H., Colloquium Lectures on Geometric Measure Theory, *Bull. Amer. Math. Soc.* **84** (1978), pp. 291–338.

[6] Harvey R., Holomorphic Chains and Their Boundaries, *Proceedings of Symposia in Pure Mathematics*, **30** : **1**, A.M.S., Providence, R.I. (1977), pp. 309–382.

[7] Harvey R., Lawson H. B., Jr., On Boundaries of Complex Analytic Varieties, I, *Ann. of Math.* **102** (1975), pp. 233–290.

[8] Harvey R., Lawson H. B., Jr., On Boundaries of Complex Analytic Varieties, II, *Ann. of Math.* **106** (1977), pp. 213–238.

[9] Harvey R., Lawson H. B., Jr., Calibrated Foliations, *Amer. J. Math.* **103** (1981), pp. 411–435.

[10] Harvey R., Lawson H. B., Jr., Calibrated Geometries, *Acta Math.* **148** (1982), pp. 47–157.

[11] Harvey R., Lawson H. B., Jr., The Cartan–Kähler Theorem Applied to Boundaries of Minimal Surfaces in Calibrated Geometries. (An unpublished supplement to [10].)

[12] Harvey R., Lawson H. B., Jr., An Intrinsic Characterization of Kähler Manifolds. *Invent. Math.* **74** (1983), pp. 169–198.

[13] Harvey R., Morgan F. *The Faces of the Grassmannian of Three Planes in* $\boldsymbol{R}^7$ (in preparation).

[14] Harvey R., Polking J., Removable Singularities of Solutions of Linear Partial Differential Equations, *Acta Math.* **125** (1970), pp. 39–56.

[15] Hsiang W. Y., Lawson H. B., Jr., Minimal Submanifolds of Low Cohomogeneity, *J. Differential Geom.* **5** (1971), pp. 1–38.

[16] Lawson H. B., Jr., The Equivariant Plateau Problem and Interior Regularity, *Trans. Amer. Math. Soc.* **173** (1972), pp. 231–250.

[17] Lawson H. B., Jr., Osserman R., Non-Existence, Non-Uniqueness and Irregularity of Solutions to the Minimal Surface System, *Acta Math.* **139** (1977), pp. 1–17.

[18] Lewy H., On the Non-Vanishing of the Jacobian of a Homeomorphisms by Harmonic Gradients, *Ann. of Math.* **88** (1968), pp. 578–629.

[19] Michelsohn M. L., On the Existence of Special Metrics in Complex Geometry (to appear).

[20] Miyaoka Y., Kähler Metrics on Elliptic Surfaces, *Proc. Japan Acad.* **50** (1974), pp. 533–536.
[21] Morgan F., On the Singular Structure of Three-Dimensional Area-Minimizing Surfaces, *Trans. A.M.S.* (to appear).
[22] Siu Y.-T., Every K3 Surface in Kähler, *Invent. Math.* (to appear).
[23] Sullivan D., Cycles for the Dynamical Study of Foliated Manifolds and Complex Manifolds, *Invent. Math.* **36** (1976), pp. 225–255.
[24] Sullivan D., A Homological Characterization of Foliations Consisting of Minimal Surfaces, *Comm. Math. Helv.* **54** (1979), pp. 218–223.
[25] Todorov A. N., Applications of the Kähler–Einstein–Calabi–Yau Metric to Moduli of K3 Surfaces, *Invent. Math.* **61** (1980), pp. 251–265.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

G. M. HENKIN

Tangent Cauchy–Riemann Equations and the Yang–Mills, Higgs and Dirac Fields

The first part of the paper contains a survey of conditions for the local and global solvability of the tangent Cauchy–Riemann equations on q-concave CR-manifolds.

In the second part results are presented concerning the representation (by the Radon–Penrose type transformation) of the classical Yang–Mills, Higgs and Dirac fields as solutions of the Cauchy–Riemann equations on 1-concave submanifolds of twistor (or supertwistor) space.

1. Cauchy–Riemann equations on q-concave CR-manifolds

1.1. $\bar{\partial}_\tau$-closed forms and their local approximation by $\bar{\partial}_\tau$-exact forms. Let X be an n-dimensional complex manifold, let E be a holomorphic vector bundle over X. We denote by $\bar{\partial}$ a Cauchy–Riemann operator annihilating all holomorphic sections of the fibre bundle E over X. Let L be a real, closed submanifold of X of co-dimension k which can be represented in each coordinate neighbourhood $\Omega \subset X$ in the form:

$$L \cap \Omega = \{z \in \Omega \colon \varrho_1(z) = \ldots = \varrho_k(z) = 0\}, \tag{1.1}$$

where $\{\varrho_\nu\}$ are smooth real-valued functions in the domain $\Omega \subset X$ satisfying the condition $\bar{\partial}\varrho_1 \wedge \ldots \wedge \bar{\partial}\varrho_k \neq 0$ on $L \cap \Omega$.

For a fixed point $p \in L$ the complex tangent space $T_p^c(L)$ has a complex dimension $n-k$ and in local coordinates $z = (z_1, \ldots, z_n)$ is determined by the equations

$$T_p^c(L) = \left\{\zeta \in C^n \colon \sum_{j=1}^{n} \frac{\partial \varrho_\nu}{\partial z_j}(p)\zeta_j = 0, \quad \nu = 1, 2, \ldots, k\right\}.$$

Such a manifold is called a (generic) CR-manifold.

Let us denote by $C^{(s)}_{0,q}(L, E)$, $0 \leqslant q \leqslant n-k$, $s \geqslant 0$, the space of differential forms of type $(0, q)$ on M with E-valued $C^{(s)}$-smooth coefficients.

If $f \in C_{0,q-1}(L, E)$ and $g \in C_{0,q}(L, E)$ are such that for any compactly supported form $\varphi \in C^{(\infty)}_{n,n-k-q}(X, E^*)$, where E^* is the fibre bundle dual to E, we have

$$\int_L g \wedge \varphi = (-1)^q \int_L f \wedge \bar{\partial}\varphi,$$

then, by definition, we shall write

$$\bar{\partial}_\tau f = g, \tag{1.2}$$

where $\bar{\partial}_\tau$ is the tangent Cauchy–Riemann operator.

A necessary condition for (1.2) to be locally solvable is, first of all, the condition $\bar{\partial}_\tau g = 0$ on L. The forms (resp. the functions) satisfying this condition are called CR-forms (resp. CR-functions).

If the manifold L and the form g are real-analytic then the condition $\bar{\partial}_\tau g = 0$ is also sufficient for the local solvability of (1.2). In this case the dimensions of the domains in M where (1.2) is solvable depend not only on the manifold L but also on real-analytic properties of the CR-form g (see [43]).

If either the form g or the manifold M is not real-analytic, then, generally speaking, the condition $\bar{\partial}_\tau g = 0$ is no longer sufficient for the local solvability of (1.2) (Hans Lewy's effect).

In this case, however, an important general result concerning the approximate local solvability of equation (1.2) is proved by M. Baouendi and F. Treves.

THEOREM 1.1 (F. Treves [45]). *Let L be a smooth CR-submanifold (of co-dimension k) in a complex manifold X. Then for any point $p \in L$ there exists a neighbourhood Ω_p such that every CR-form $g \in C^{(s)}_{0,r}(L)$, $r = 0, 1, \ldots$ $\ldots, n-k$, can be approximated on $L \cap \Omega_p$ by $\bar{\partial}$-closed forms from $C^{(\infty)}_{0,r}(\Omega_p)$ as exactly as desired in $C^{(s)}$-topology. If, moreover, $r > 0$ then these forms are $\bar{\partial}_\tau$-exact on $L \cap \Omega_p$.*

For hypersurfaces this result was proved earlier (see [1, 13]).

1.2. q-concave CR-manifolds and the local exactness of CR-forms of type $(0, r)$ for $r < q$ and $\tau > n-k-q$. The study of conditions for the exact (and not merely approximate) local and global solvability of the equation (1.2), initiated by H. Lewy in his classical work, has been carried sufficiently far for the case where L is a hypersurface in X (see J. J. Kohn [25],

J. Kohn, H. Rossi [26], A. Andreotti, C. D. Hill [1], G. M. Henkin [13], A. Bogges [6]).

In recent years the results of these works have been generalized to the case of CR-manifolds of arbitrary co-dimension. The formulations of the main results use E. Levi's form of a manifold L. The last one is given by the equality

$$L_{p,\lambda}(L)(\zeta) = \sum_{\nu=1}^{k} \sum_{\alpha,\beta=1}^{n} \lambda_\nu \left(\frac{\partial^2 \varrho_\nu}{\partial z_\alpha \partial \bar{z}_\beta} (p) \zeta_\alpha \bar{\zeta}_\beta \right),$$

where $p \in L$, $\zeta \in T_p^c(L)$, $\lambda \in R^k$.

The manifold L is said to be q-concave (resp. weak q-concave) at the point $p \in L$, if for all $\lambda \in R^k \setminus \{0\}$ the form $L_{p,\lambda}(L)$ has on $T_p^c(L)$ at most q negative (resp. q nonpositive) eigenvalues.

The basic theorems on the local solvability of the equation (1.2) result from the following general proposition on the $\bar{\partial}$-closed extension of CR-forms into a neighbourhood of a generic CR-manifold.

THEOREM 1.2 (G. M. Henkin [16]). *If the CR-manifold of the form* (1.1) *is q-concave then there exists a neighbourhood X' of the manifold L, such that for all r satisfying condition $0 \leqslant r < q$ or $n-k-q < r \leqslant n-k$, and for any CR-form $f \in C_{0,r}^{(s)}(L, E)$, $s \geqslant 0$, there exists a $\bar{\partial}$-closed form $F \in C_{0,r}^{(s-1/2-\varepsilon)}(X', E)$ such that $F|_L = f$ and $F \wedge \partial\varrho_1 \wedge \ldots \wedge \bar{\partial}\varrho_k \in C^{(s)}(\Omega \cap X')$ for every coordinate neighbourhood Ω on X.*

The method used in the proof of Theorem 1.2 actually gives an explicit integral formula for F in terms of f.

In the case of CR-functions Theorem 1.2 confirms a conjecture due to I. Naruki [35], where the respective statement concerning CR-functions was proved for "standard" CR-manifolds, i.e. for manifolds of the type

$$L = \{(z, w) \in C^k \times C^{n-k} \colon \operatorname{Im} z_\nu = F_\nu(w, \bar{w}),\ \nu = 1, 2, \ldots, k\},$$

where $\{F_\nu\}$ are Hermitian forms on C^{n-k}.

For hypersurfaces and $s = \infty$ Theorem 1.2 was first obtained in a paper by A. Andreotti and C. D. Hill [1].

From Theorem 1.2 the following facts can be deduced concerning the local solvability of the equation (1.2).

THEOREM 1.2a ([16]). *Under the conditions of Theorem 1.2 for any point $p \in L$ and any sufficiently small neighbourhood Ω_p of the point p, for all r: $1 \leqslant r < q$, and for any CR-form $g \in C_{0,r}^{(s)}(L, E)$, there exists a form $f \in C_{0,r-1}^{(s+1/2-\varepsilon)}(L \cap \Omega_p, E)$ satisfying on $L \cap \Omega_p$ the equality $\bar{\partial}_\tau f = g$.*

Theorem 1.2a was regarded in the literature as a plausible conjecture (see [2, 44]). Earlier similar results had been obtained (for $s = \infty$) in the work of F. Treves [44] for "tubelike" CR-manifolds, i.e., for manifolds of the type

$$L = \{z \in C^n\colon \operatorname{Im} z_\nu = \varphi_\nu(\operatorname{Im} z_{k+1}, \ldots, \operatorname{Im} z_n),\ \nu = 1, 2, \ldots, k\},$$

and also for "standard" manifolds, in the work of H. Rossi, M. Vergne [40]. Moreover, in the work of M. Sato, T. Kawai, M. Kashivara [41], the microlocal variant of Theorem 1.2a was established.

THEOREM 1.2b ([16]). *Under the conditions of Theorem* 1.2, *for any Stein domain $\Omega \subset X$, for all r, $n-k-q < r \leqslant n-k$, and for any CR-form $g \in C^{(s)}_{0,r}(L\cap\Omega, E)$, there exists a form $f \in C^{(s+1/2-\varepsilon)}_{0,r-1}(L\cap\Omega, E)$ satisfying on $L\cap\Omega$ the equality $\bar\partial_\tau f = g$.*

For $s = \infty$ Theorem 1.2b was first obtained in a very important paper by I. Naruki [34].

1.3. A criterion for the local solvability of "nonsolvable" tangent Cauchy–Riemann equations in q-concave manifolds. The requirement of q-concavity of the CR-manifold L in the formulation of Theorem 1.2 would be an adequate condition provided that the following result explaining H. Lewy's effect were true.

THEOREM 1.3 (A. Andreotti, G. Fredricks, M. Nacinovich [2]). *If for a CR-manifold of type* (1.1), *for some $p \in L$ and $\lambda \in R^k\setminus\{0\}$, the form $L_{p,\lambda}(L)$ is not degenerate on $T^c_p(L)$ and has q negative and $n-k-q$ positive eigenvalues, then for any sufficiently small neighbourhood U of the point p there exists a CR-form f from $C^{(\infty)}_{0,q}(L\cap U, E)$ which is not $\bar\partial_\tau$-exact on $L\cap U$.*

A microlocal variant of Theorem 1.3 was obtained earlier in a paper by M. Sato, T. Kawai, M. Kashivara [41].

For hypersurfaces Theorem 1.3 was obtained earlier in a paper by A. Andreotti and C. Hill [1].

Theorem 1.3 shows the necessity of complementary conditions for the solvability of the equation (1.2) when g is a CR-form of type $(0, q)$ on a q-concave CR-manifold. A criterion for the local solvability of (1.2) in this case was obtained in [16]. This criterion we shall now formulate for real-analytic CR-manifolds only, in the form of a criterion of extendability of the CR-form g to a $\bar\partial$-closed form $\tilde g$ in a neighbourhood of L.

THEOREM 1.4 ([16]). *Let L be a real-analytic q-concave CR-manifold of type* (1.1). *A CR-form $g \in C_{0,q}(L, E)$ can be extended to a $\bar{\partial}$-closed form $\tilde{g}$ in some neighbourhood of L if and only if, given any $p \in L$, there exists a neighbourhood Ω_p such that the $(0, q)$-form described by:*

$$Kf(z) = \int_{\zeta \in \Omega_p \cap L} f(\zeta) K_q(\zeta, z)$$

is real-analytic on $L \cap \Omega_p$, where $K_q(\zeta, z)$ is a suitable singular kernel of the Cauchy–Fantappe type which is a real-analytic CR-form of type $(n, n-k-q)$ with respect to the variable $\zeta \in (\Omega_p \cap L) \setminus \{z\}$.

For hypersurfaces a corresponding criterion was established earlier (see [13]).

Theorem 1.4 is a basis for the proof of the following more subtle solvability criterion, which we shall formulate as the "edge of wedge" theorem or CR-forms.

THEOREM 1.4a (R. A. Ayrapetian, G. M. Henkin [4]). *Let a q-concave fgeneric CR-manifold L of co-dimension k in X belong to CR-manifolds L_j, $j = 1, 2, \ldots, k$ of co-dimension $k-1$ such that for all $p \in L$ the tangent spaces $T_p(L)$ belong to the linear hull of the complex tangent spaces $\{T_p^c(L_j)\}$. Then the CR-form g from $C_{0,q}(L)$ admits a $\bar{\partial}$-closed extension into a neighbourhood of the manifold L if the form g is a trace on L of some CR-form $\tilde{g} \in C_{0,q}(\bigcup_{j=1}^{k} L_j)$.*

For $q = 0$ Theorem 1.4a is a generalization of a number of results extending the classical theorems of S. N. Bernstein on separate analyticity and of N. N. Bogolubov on the "edge of wedge" (see [5], [46]).

It is also Theorem 1.4 on which is based the proof of the following result concerning the solvability of the equation (1.2) on a q-concave manifold for $(0, q)$-forms with a (sufficiently) compact support.

THEOREM 1.4b (G. M. Henkin [19]). *Given a q-concave CR-manifold L with $q \geqslant 1$, for any point $p \in L$, any sufficiently small pseudoconvex neighbourhood Ω_p of this point, any r: $1 \leqslant r \leqslant q$, and any CR-form $g \in C_{0,r}^{(s)}(L, E)$ with a support in $\Omega_p \cap L$, there exists a form $f \in C_{0,r-1}^{(s+1/2-\varepsilon)}(L, E)$ with a support in $\Omega_p \cap L$ satisfying on L the equation $\bar{\partial}_\tau f = g$.*

For $r < q$ (and for $s = \infty$) Theorem 1.4b was obtained by I. Naruki [34]. Theorem 1.4b seems to be a new one even for hypersurfaces. In particular it implies that for a CR-function on any 1-concave CR-manifold L the

Hartoggs–Bochner effect holds. This generalizes N. Sibony's earlier statement [42] concerning the validity of the local maximum principle for CR-functions on a 1-concave CR-manifold.

1.4. Conditions for a global solvability of the tangent Cauchy–Riemann equations. Let us denote by $A^{(s)}_{0,q}(L, E)$ the subspace in $C^{(s)}_{0,q}(L, E)$ composed of $\bar\partial_\tau$-closed forms, and by $B^{(s)}_{0,q}(L, E)$ the subspace in $C^{(s)}_{0,q}(L, E)$ composed of forms $\bar\partial_\tau f$, where $f \in C^{(s)}_{0,q-1}(L, E)$. We shall consider the space of $\bar\partial_\tau$-cohomologies

$$H^q_{(s)}(L, E) = A^{(s)}_{0,q}(L, E)/B^{(s)}_{0,q}(L, E).$$

Theorems 1.2 and 1.4 enable us to establish a criterion for the global solvability of equation (1.2) together with a criterion for the finitely dimensionality of the cohomology space $H^{(q)}_{(s)}(L, E)$.

THEOREM 1.5 ([16]). *Under the conditions of Theorem 1.2 for any r: $1 \leqslant r \leqslant q$ (resp. $n-k-q < r \leqslant n-k$) and any pseudo-concave (resp. pseudo-convex) domain $\Omega \Subset X$, a necessary and sufficient condition for a CR-form f from $A^{(s)}_{0,r}(L\cap\Omega, E)$ to be $\bar\partial_\tau$-exact on $L\cap\Omega$, and to belong to the space $B^{(s)}_{0,r}(L\cap\Omega, E)$ is that $\int\limits_L f\wedge\varphi = 0$ for any form $\varphi \in A^{(\infty)}_{n,n-k-r}(X, E^*)$ with a support in the domain Ω. Furthermore, for all $r < q$ (resp. $r > n-k-q$), the spaces $H^r_{(s)}(L\cap\Omega, E)$ are finitely dimensional.*

In the case where L is a compact hypersurface in X, the statement of Theorem 1.5 has been well known (see J. Kohn, H. Rossi [25], [26]).

With regard to the subjects considered in the second part of the paper we shall now take into consideration smooth $(0, 1)$-forms θ on L with values in End E. Such a form we shall call $\bar\partial_\tau$-exact if

$$K^{-1}\,\partial_\tau K = \theta \quad \text{on } L, \tag{1.3}$$

where K is a smooth function with values in non-degenerate endomorphisms E.

A necessary condition for the local solvability of (1.3) is now the equality:

$$\bar\partial_\tau\theta + \theta\wedge\theta = 0. \tag{1.4}$$

The space of the smooth (of the class C^∞) forms θ satisfying (1.4) and considered up to the transformation of the form

$$\theta \sim \tilde\theta = K^{-1}\,\bar\partial_\tau K + K^{-1}\theta K$$

we shall further denote by $H^{0,1}(L, \mathrm{GL}(E))$.

Any element θ from $H^{0,1}(L, \mathrm{GL}(E))$ defines a CR-fibre bundle E_θ over L which is topologically equivalent to the fibre bundle E. Namely, those smooth sections h of the fibre bundle E which are annihilated by the operator $\bar{\partial}_\tau + \theta$,

$$\bar{\partial}_\tau h + \theta h = 0,$$

we shall call CR-sections of the fibre bundle E_θ.

In contradistinction to holomorphic fibre bundles over X, a CR-fibre bundle over L cannot always be given by transition matrix-functions. For instance, from Theorem 1.2 it follows that on a 1-concave CR-manifold the CR-fibre bundle E can be given by means of CR-matrix-functions of pass on some cover if and only if E_θ is a trace on L of some holomorphic fibre bundle in a neighbourhood of L. As above, using the Cauchy–Riemann operator $\bar{\partial}_\tau + \theta$, cohomology spaces $H^q_{(s)}(L, E_\theta)$ are introduced. For a q-concave CR-manifold L, for any $\theta \in H^q_{(s)}(L, \mathrm{GL}(E))$ and any r: $1 \leqslant r < q$ or $r > n-k-q$, the local results of Theorems 1.2b and 1.4b and the global result of Theorem 1.5 are still valid.

2. Yang–Mills, Higgs and Dirac fields as solutions of Cauchy–Riemann equations

R. Penrose ([36], [37]) has proposed a promising program of a reconstruction of the foundations of relativistic physics, which would result in a transformation of relativistic physics into a part of analytic geometry in the space of complex light lines (the theory of twistors).

We shall present a number of results developing Penrose's program. These results prove that the theory of the classical Yang–Mills, Higgs and Dirac fields on Minkowski spaces can be transformed into the theory of Cauchy–Riemann tangent equations on a 1-concave submanifold of the twistor (supertwistor) space. Some of the results of the first part of the paper acquire here a "physical" interpretation.

2.1. The spaces of complex and real zero lines. Let CM_0 be a complex Minkowski space, i.e., a four-dimensional complex space with spinor coordinates $u = \{u_{AB'}, A = 0, 1;\ B' = 0', 1'\}$ and with a metric $\det(du_{AB'})$. Then a real Minkowski space M_0 is formed of those points $u \in CM_0$ for which the matrix $u_{AB'}$ is Hermitian.

Moreover, let T_+ and T_- be two reciprocally dual four-dimensional complex spaces (of twistors and dual twistors) with coordinates $(z^A, z_{B'})$

and $(w_A, w^{B'})$, and with the bilinear form $\langle z\cdot w\rangle = z^A w_A - z_{A'} w^{A'}$; then the coordinates in T_+ and T_- are the homogeneous coordinates of the corresponding points in the three-dimensional projective spaces $\mathscr{L}_\pm = P(T_\pm)$.

The equation $\langle z\cdot w\rangle = 0$ distinguishes a five-dimensional hypersurface $\mathscr{L}$ on $\mathscr{L}_+\otimes\mathscr{L}_-$. The points of this surface parametrize complex zero-lines in the Minkowski space; given a fixed point $(z, w)\in\mathscr{L}$, there exists a corresponding zero-line in CM_0, given by the equations:

$$z_{B'} = u_{AB'} z^A, \quad w_A = u_{AB'} w^{B'}. \tag{2.1}$$

Conversely, for a fixed point $u\in CM_0$, the equations (2.1) considered as conditions for (z, w), generate two reciprocally orthogonal two-dimensional subspaces in T_+ and T_-, which, on the other hand, determine on $\mathscr{L}$ a quadric $\mathscr{L}(u) = \mathscr{L}_+(u)\otimes\mathscr{L}_-(u)$.

The foregoing correspondence enables us to identify the compactifiable (and complexifiable) Minkowski space CM with the manifold of all two-dimensional subspaces in T_+ (or in T_-). Corresponding tautological two-dimensional fibre bundles over CM, denoted by $S_\pm$, are called spinor fibre bundles.

In virtue of (2.1), over CM_0 there are natural trivializations of these fibre bundles; z^A are the coordinates in the fibre $S_+(u)$, $w^{A'}$ are the coordinates in the fibre $S_-(u)$, $u\in CM_0$.

Given a domain U in a compactifiable real Minkowski space M, we shall denote by $L(U)$, $L_+(U)$, $L_-(U)$ the real submanifolds in the complex manifolds $\mathscr{L}, \mathscr{L}_\pm$, of the form:

$$L(U) = \bigcup_{u\in U}\mathscr{L}(u), \quad L_\pm(U) = \bigcup_{u\in U}\mathscr{L}_\pm(u).$$

Let us put $L = L(M)$, $L_\pm = L_\pm(M)$. Each of the manifolds L_+ and L_- parametrizes the (real) light rays on the real Minkowski space M, whereas the manoifold L parametrizes pairs of intersecting world lines on M (or, in other words, the complex zero-lines in CM, passing through M). We have

$$\begin{aligned} L_+ &= \{z\in\mathscr{L}_+ : \operatorname{Im}(z^0\bar z_{0'} + z^1\bar z_{1'}) = 0\}, \\ L_- &= \{w\in\mathscr{L}_- : \operatorname{Im}(w_0\overline{w}^{0'} + w_1\overline{w}^{1'}) = 0\}, \\ L &= \{(z, w)\in\mathscr{L} : z\in L_+, w\in L^-\}. \end{aligned}$$

The manifolds $L_\pm$ are real hypersurfaces in the spaces $\mathscr{L}_\pm$, whose Levi forms are non-degenerate at any point $p\in L_\pm$ and have one positive and one negative eigenvalue on $T_p^c(L_\pm)$, i.e., $L_\pm$ are strongly 1-concave manifolds.

The manifold L, of co-dimension 2 in the complex manifold $\mathscr{L}$, has singularities on a fully real submanifold

$$S = \{(z, w) \in L\colon (w_A, w^{B'}) = (\bar{z}^{A'}, \bar{z}_B)\}.$$

For all $p \in L \setminus S$ we have $\dim T_p^c(L) = 3$, and for any $\lambda \in R^2 \setminus \{0\}$ the Levi form $L_{p,\lambda}(L)$ has on $T_p^c(L)$ one positive, one negative, and one zero eigenvalue, i.e., L is a 1-concave and, at the same time, weak 2-concave CR-manifold outside S.

Further, let E be a trivial n-dimensional fibre bundle. We shall also assume that the domain $U \subset M$ is such that its intersections with all light rays are connected and simply connected. Further, let us also denote by $H^{0,1}(L(U), \mathrm{GL}(n, C))$ (resp. $H^{0,1}(L_\pm(U), \mathrm{GL}(n, C))$) a set of all CR-fibre bundles E_θ topologically equivalent to E and moreover analytically equivalent to E on each quadric $L(u)$, $u \in U$. The last statement means that the $(0, 1)$-form indexing the fibre bundle E_θ can be represented on each quadric $L(u)$

$$\theta|_{L(u)} = K_u^{-1} \bar{\partial} K_u, \tag{2.2}$$

where the function K_u with values in $\mathrm{GL}(n, C)$ smoothly depends on the parameter $u \in U$.

For one-dimensional topologically trivial fibre bundles for instance, the condition (2.2) holds automatically.

In the sequel we shall put $E_\theta(m, k) = E_\theta \otimes \mathcal{O}(m, k)$, where $\mathcal{O}(m, k)$ is a one-dimensional fibre bundle over $\mathscr{L}_+ \otimes \mathscr{L}_-$, the holomorphic sections of which are holomorphic functions on $T_+ \otimes T_-$ of homogeneity (m, k) with respect to variables $(z, w) \in T_+ \otimes T_-$.

Theorems 1.2, 1.3, 1.5 applied to the CR-manifolds $L(U)$, $L_+(U)$, $L_-(U)$ enable us to state, first, that the spaces of CR-fibre bundles $H^{0,1}(L(U), \mathrm{GL}(n, C))$ and $H^{0,1}(L_\pm(U), \mathrm{GL}(n, C))$ are non-trivial (and infinitely dimensional) and, secondly, that among spaces of cohomologies with coefficients in the fibre bundles $E_\theta(m, k)$ there are in general no other non-trivial (infinitely dimensional) spaces but the spaces $H^1(L_\pm(U), E_\theta(m))$, $H^1(L(U), E_\theta(m, k))$ and $H^2(L(U), E_\theta(m, k))$.

R. Penrose's transformation enables us to identify, in a surprisingly natural way, the elements of these spaces as cohomologies, and also the fibre bundles E_θ as physical fields on Minkowski spaces.

2.2. The Penrose transformation and criteria of solvability of the tangent Cauchy–Riemann equations on $L(U)$. Let a CR-fibre bundle E_θ be indexed by the elements $\theta \in H^{0,1}(L(U),\ \mathrm{GL}(n, C))$, and let $K = K(u, z^A, w^{B'})$ be a function satisfying (2.2). Let $d_1 = z^A w^{B'}\, \partial/\partial u^{AB'}$ be an operator of differentiation along light lines. (1.2) implies that the function $(d_1 K)K^{-1}$ is holomorphic on each quadric $L(u)$, $u \in U$, and therefore can be represented in the form

$$(d_1 K)K^{-1} = z^A w^{B'} a_{AB'}(u), \tag{2.3}$$

where $\{a_{AB'}\}$ are smooth functions of $u \in U$.

We shall consider a 1-form $a = a_{AB'}du^{AB'}$ in the domain $U \subset M$. It is determined by the given construction up to the gauge $a \sim \tilde{a} = b^{-1}db + b^{-1}ab$, where b is a $\mathrm{GL}(n, C)$-valued smooth function. Thus the form a determines a $\mathrm{GL}(n, C)$-connection ∇_a in the fibre bundle E over $U \subset M$. The correspondence $\theta \mapsto \mathscr{P}\theta = \nabla_a$ will be called the Penrose transform of the form θ.

For the elements of cohomology spaces with the coefficients in the fibre bundles $E_\theta(m, k)$ the definition of the Penrose transforms depends rather strongly on the numbers m and k. In the cases which are most interesting for us these definitions are the following: If a fixed function K_u satisfies (2.2), or, in other words, a form a satisfies (2.3), then, for $\Psi_+ \in H^1(L(U),\ E_\theta(-1, 0))$, $\varphi_+ \in H^1(L(U),\ E_\theta(-2, 0))$, and $\Omega \in H^1(L(U),$ $\mathrm{End}\, E_\theta(-1, -1))$ we shall put

$$\psi^{A'} = \mathscr{P}\Psi_+ = \int\limits_{L_+(u)} K \frac{\partial}{\partial z_{A'}} \Psi_+ \wedge z^A dz_A, \qquad \Phi_+ = \mathscr{P}\varphi_+ = \int\limits_{L_+(u)} K\varphi \wedge z^A dz_A, \tag{2.4}$$

$$\omega = \mathscr{P}\Omega = \int\limits_{L_+(u)} K\left(w^{A'} \frac{\partial}{\partial z^{A'}} \Omega\right) K^{-1} \wedge z^A dz_A. \tag{2.4^1}$$

For $F_+ \in H^2(L(U),\ E_\theta(-3, -1))$, $G_+ \in H^2(L(U),\ E_\theta(-3, -2))$ and $J \in H^2(L(U),\ \mathrm{End}\, E_\theta(-3, -3))$ we shall put

$$f_+ = \mathscr{P}F_+ = \int\limits_{L(u)} K z^A \frac{\partial F_+}{\partial w^A} \wedge z^B dz_B \wedge w^{B'} dw_{B'},$$

$$g_A = \mathscr{P}G_+ = \int\limits_{L(u)} z_A K G_+ \wedge z^B dz_B \wedge w^{B'} dw_{B'}, \tag{2.5}$$

$$j = \mathscr{P}J = j_A^{B'} du^{A0'} \wedge du^{A1'} \wedge du_{AB'}, \tag{2.5^1}$$

where

$$j_A^{B'} = \iint_{L(u)} KJK^{-1} z_A w^{B'} \wedge z^B dz_B \wedge w^{A'} dw_{A'}.$$

The Penrose transforms for cohomologies with coefficients in $E_\theta(0, -1)$, $E_\theta(0, -2)$, $E_\theta(-1, -3)$, $E_\theta(-2, -3)$ are defined analogously. Let us note that the fields $\psi^{A'}$, φ_+, etc. are determined by the given construction up to the gauges $\psi^{A'} \sim \tilde{\psi}^{A'} = b\psi^{A'}$; $\varphi_+ \sim \tilde{\varphi}_+ = b\varphi_+$ etc., where b is a gauge function for the connection a.

The following result sums up (and to a certain extent generalizes) a number of statements given in [47], [9], [12], [49], [24], [30], [14], [20], [21], [7].

THEOREM 2.1. *The Penrose transformation of the form* (2.3) *establishes a canonical isomorphism between the space of the CR-fibre bundles* $\theta \in H^{0,1}(L(U), \mathrm{GL}(n, C))$ *and the space of all smooth connections* $\nabla_a = \mathscr{P}\theta$ *in the fibre bundle* E *over* U. *Moreover, the CR-fibre bundles from* $H^{0,1}(L_\pm(U), \mathrm{GL}(n, C))$ *are transformed into self-dual (resp. anti-self-dual) connections.*

Subsequently, for fixed $\theta \in H^{0,1}(L(U), \mathrm{GL}(n, C))$, *the Penrose transformations of the form* (2.4), (2.5) *establish an isomorphism of the cohomology spaces* $H^1(L(U), E_\theta(-1, 0))$, $H^1(L(U), E_\theta(-2, 0))$, $H^2(L(U), E_\theta(-3, -2))$, $H^2(L(U), E_\theta(-3, -1))$ *with the spaces of smooth sections over* U *of the fibre bundles* $E \otimes S_- \otimes \Lambda^2 S_+$; $E \otimes \Lambda^2 S_+$; $E \otimes S_+ \otimes \Lambda^2 S_+ \otimes \Lambda^2 S_-$; $E \otimes (\Lambda^2 S_+)^2 \otimes \Lambda^2 S_-$, *respectively. If, moreover,* $\theta \in H^{0,1}(L_+(U), \mathrm{GL}(n, C))$ *then the elements of the spaces* $H^1(L_+(U), E_\theta(-1, 0))$ *and* $H^1(L_+(U), E_\theta(-2, 0))$ *are transformed into the solutions* $\psi_{A'}$ *and* φ_+ *of the Weil–Dirac equation* $\nabla_a^{AA'} \psi_{A'} = 0$ *and the d'Alembert equation* $\Box_a \varphi_+ = 0$, *respectively, in the self-dual field* $a = \mathscr{P}\theta$.

Finally, the correspondence of (2.4¹) *and* (2.5¹) *realizes the isomorphisms of the spaces* $H^1(L(U), \mathrm{End}\, E_\theta(-1, -1))$ *and* $H^2(L(U), \mathrm{End}\, E_\theta(-3, -3))$ *with, respectively, the space of smooth sections of the fibre bundle* $\mathrm{End}\, E \otimes \Lambda^2 S_+ \otimes \Lambda^2 S_-$ *and the space of smooth 3-forms* j *on* U *with values in* $\mathrm{End}\, E$, *satisfying the equation* $dj + [a, j] = 0$.

From Theorems 2.1 and 1.2, as a simple corollary, follows

THEOREM 2.1a. *Let the* $(0,1)$*-forms* Θ, $\Phi_\pm$, $\Psi_\pm$ *represent elements of the respective spaces of one-dimensional cohomologies on* $L(U)$ *(or* $L_\pm(U)$*). Then a necessary and sufficient condition for the Cauchy–Riemann equations of the forms* $a^{-1}\bar{\partial}_\tau a = \theta$; $\bar{\partial}_\tau \beta_\pm + \theta\beta_\pm = \Phi_\pm$; $\bar{\partial}_\tau \gamma_\pm + \theta\gamma_\pm = \Psi_\pm$ *to be solvable (resp. locally solvable in a neighbourhood of any point) is that the Penrose* *transforms $\mathscr{P}\theta$, $\mathscr{P}\Phi_{\pm}$, $\mathscr{P}\Psi_{\pm}$ of these forms be equal to zero (resp. real-analytic) on U.*

Let us note that a following criterion for the $\bar{\partial}_\tau$-exactness of the forms $\Phi_{\pm}$, $\Psi_{\pm}$ in a neighbourhood of a fixed point $p \in L_{\pm}(U)$ was formulated in [17] as a development of Penrose's ideas [38]: the singular spectrum of the forms $\mathscr{P}\Phi_{\pm}$, $\mathscr{P}\Psi_{\pm}$ must not contain zero-bi-characteristics corresponding to light rays $l_{\pm}(\tau)$.

2.3. The Maxwell–Yang–Mills, Weil–Dirac and Klein–Gordon equations as the Cauchy–Riemann equations. Let $L^{(j)}(U)$ denote the j-th infinitesimal neighbourhood of the manifold $L(U) \subset L_+(U) \times L_-(U)$. We shall denote by $H^{0,1}(L^{(j)}(U), \mathrm{GL}(n, C))$ the subspace of CR-fibre bundles in $H^{0,1}(L(U), \mathrm{GL}(n, C))$ indexed by smooth $(0,1)$-forms θ on $L_+(U) \times L_-(U)$ which, first, satisfy on $L^{(j)}(U)$ the Cauchy–Riemann–Cartan equation of the form $\bar{\partial}\theta + \theta \wedge \theta = X_{j+1}\langle z \cdot w\rangle^{j+1}$ where X_{j+1} is a smooth $(0,2)$-form on $L(U)$ representing some element of the space $H^2(L(U), \mathrm{End}\,E_\theta(-j-1, -j-1))$, and, secondly, are considered up to the $\bar{\partial}$-gauge of the form

$$\theta \sim \tilde{\theta} = K^{-1}\bar{\partial}K + K^{-1}\theta K + O(\langle z \cdot w\rangle^{j+1}).$$

We shall denote by $H^q(L^{(j)}(U), E_\theta)$, where $\theta \in H^{0,1}(L^{(j)}(U), \mathrm{GL}(n, C))$, a subspace in $H^q(L(U), E_\theta)$ given by smooth E-valued $(0,q)$-forms ω on $L_+ \times L_-$ satisfying on $L^{(j)}(U)$ the Cauchy–Riemann equation $\bar{\partial}_\tau\omega + \theta \wedge \omega = Y_{j+1}\langle z \cdot w\rangle^{j+1}$ and considered up to the $\bar{\partial}$-gauge $\omega \sim \tilde{\omega} = \omega + \bar{\partial}_\tau a + \theta \wedge a + O(\langle z \cdot w\rangle^{j+1})$.

It was proved in [14], [20] that any element of the space $H^{0,1}(L(U), \mathrm{GL}(n, C))$ can be extended (in only one way) to an element of the space $H^{0,1}(L^{(2)}(U), \mathrm{GL}(n, C))$ and can be represented by the form θ on $L_+(U) \times L_-(U)$ satisfying the relation

$$\bar{\partial}_\tau\theta + \theta \wedge \theta = J\langle z \cdot w\rangle^3. \tag{2.6}$$

Moreover, the elements of the space $H^{0,1}(L^{(1)}(U), \mathrm{GL}(n, C))$ can be indexed by the forms $\tilde{\theta}$

$$\tilde{\theta} = \theta + \Omega\langle z \cdot w\rangle, \tag{2.7}$$

where θ satisfies (2.6) and $\Omega \in H^1(L(U), \mathrm{End}\,E_\theta(-1, -1))$.

Furthermore (see [20], [21]), for fixed θ satisfying (2.6), the elements Ψ_+, Φ_+, Ω of the spaces $H^1(L(U), E_\theta(-1, 0))$, $H^1(L(U), E_\theta(-2, 0))$ and $H^1(L(U), \mathrm{End}\,E_\theta(-1, -1))$ with a suitable gauge satisfy the relations of the form

$$\begin{aligned} \bar{\partial}\Psi_+ + \theta\wedge\Psi_+ &= G_+\langle z\cdot w\rangle^2, \\ \bar{\partial}\Phi_+ + \theta\wedge\Phi_+ &= F_+\langle z\cdot w\rangle. \end{aligned} \tag{2.8}$$

The Cauchy-Riemann equations of the forms (2.6), (2.8) proved to be equivalent (on a twistor space) to the Maxwell–Yang–Mills, Weil–Dirac and Klein–Gordon equations, respectively.

THEOREM 2.2 (G. M. Henkin, Yu. Manin [14], [20], [30], [32]). *In order that the forms θ and J; Ψ_+ and G_+; Φ_+ and F_+ satisfy the equations* (2.6) *and* (2.8) *it is necessary and sufficient that their Penrose transforms of the forms* (2.3)–(2.5) *satisfy the equations*

$$d*f+[a, *f] = \frac{3}{\pi^2}j, \tag{2.6^1}$$

$$\nabla_{AA'}\psi^{A'} = \frac{1}{2\pi^2}g_A, \quad \Box\varphi_+ = \frac{1}{\pi i}f_+, \tag{2.8^1}$$

respectively, where $f = da + a\wedge a$ is the curvature form of the connection a, and $*$ is the Hodge operator corresponding to the Minkowski metric,

$$\nabla_{AA'} = \frac{\partial}{\partial u^{AA'}} + a^{AA'}, \quad \Box = \nabla^{AB'}\nabla_{AB'}.$$

From Theorems 2.2 and 2.1, as a corollary, we obtain the following result.

THEOREM 2.3 (E. Witten [50], T. Isenberg, Ph. Yasskin, P. Green [24], G. M. Henkin, Yu. I. Manin [14], [20], [32]). *The Penrose transformation establishes a canonical isomorphism between*:

(a) *the space of fibre bundles $H^{0,1}(L^{(3)}(U), \mathrm{GL}(n, C))$ and the space of all smooth $\mathrm{GL}(n, C)$-connections in E satisfying the Yang–Mills equation*

$$d*f+[a, *f] = 0,$$

(b) *the space $H^1(L^{(2)}(U), E_0)$, where $\theta\in H^{0,1}(L^{(2)}(U), \mathrm{GL}(n, C))$ and the space of the smooth solutions on U of the Weil–Dirac equation*:

$$\nabla_{AA'}\psi^{A'} = 0,$$

(c) *the space $H^1(L^{(1)}(U), E_{\tilde{0}})$, where $\theta = (\theta+\Omega\langle z\cdot w\rangle)\in H^{0,1}(L^{(1)}(U), \mathrm{GL}(n, C))$ and the space of solutions of the Klein–Gordon equation*

$$\Box\varphi_+ + \omega\varphi_+ = 0,$$

where $\omega = \mathscr{P}\Omega$.

Theorem 2.4a (for analytic connections a) was obtained by E. Witten [50] and by J. Isenberg and P. Green [24]. Theorems 2.4b and 2.4c (also for analytic fields a, $\psi_{A'}$, φ_+) were obtained by Yu. I. Manin and the present author [20], [32]. A generalization to the case of non-analytic fields a, $\psi_{A'}$ and φ_+ was obtained in [14], [15].

2.4. Super-symmetrical Yang–Mills equations and the tangent Cauchy–Riemann equations on the space of super-light rays. Physicists deal mostly with interacting Yang–Mills, Dirac or Klein–Gordon (Higgs) fields, and not with free ones. Super-symmetrical interactions of these fields (so called N-super-symmetrical Yang–Mills fields, $N = 1, 2, 3, 4$) are especially popular at present. Not entering into details, let us note that besides the connection field a (the Yang–Mills field) these equations contain: for $N = 1$, two spinor fields ψ_A and $\psi_{A'}$ (the Dirac fields); for $N = 2$, two scalar fields φ_+ and φ_- (the Higgs fields) and four spinor fields $\psi_A, \psi_{A'}, \chi_A, \chi_{A'}$; for $N = 3, 4$, six scalar fields and eight spinor fields. Moreover, the spinor fields take values in $\operatorname{End} E \otimes S_\mp \otimes \Lambda^2 S_\pm \otimes \Lambda_1$ and the scalar fields take values in $\operatorname{End} E \otimes \Lambda^2 S_\pm \otimes \Lambda_0$, where Λ_0, Λ_1 are the subspaces of respectively, even and odd elements of the Grassmann algebra Λ. The equations of motion in the 2-super-symmetrical Yang–Mills theory take the form

$$\begin{gathered}
\square\, \varphi_\pm + \{\psi_{A^\pm}, \chi^{A^\pm}\} \pm \tfrac{1}{2}\big[[\varphi_+, \varphi_-], \varphi_\pm\big] = 0, \\
\nabla_{A^-A^+} \psi^{A^\pm} \pm \tfrac{1}{2}[\varphi_\pm, \psi_{A^\mp}] = 0, \\
\nabla_{A^-A^+} \chi^{A^\pm} \pm \tfrac{1}{2}[\varphi_\pm, \chi_{A^\mp}] = 0, \\
2\nabla_{A^-A^+} f^{A^+B^+} + \{\psi_{A^-}, \chi^{B^+}\} + \\
+ \{\chi_{A^-}, \psi^{B^+}\} + \tfrac{1}{2}\big([\nabla^{B^+}_{A^-}\varphi_+, \varphi_-] - [\varphi_+, \nabla^{B^+}_{A^-}\varphi_-]\big) = 0,
\end{gathered} \tag{2.9}$$

where the symbol $A^\pm$ denotes A' and A, respectively; $\nabla_{AA'} = \dfrac{\partial}{\partial u^{AA'}} + [a^{AA'}, \cdot]$; $[\cdot, \cdot]$ and $\{\cdot, \cdot\}$ are the symbols of the commutator and the anti-commutator.

On the basis of Theorem 2.1, for any smooth fields a, $\psi_+, \psi_{A'}, \chi_A, \chi_{A'}, \varphi_\pm$ defined on $U \subset M$, one can find uniquely determined (up to $\bar\partial_\tau$-exact forms) smooth forms $\theta, \Psi_\pm, X_\pm, \Phi_\pm, \Omega$, defined on $L(U)$ with values, respectively, in $\operatorname{End} E \otimes \mathcal{O}(k, l) \otimes \Lambda_i$, where $i = 0, 1$; $(k, l) = (0, 0)$, $(-1, 0)$, $(0, -1)$, $(-2, 0)$, $(0, -2)$, $(-1, -1)$, such that, first, the Cauchy-Riemann relations of the form

$$\bar\partial_\tau\theta+\theta\wedge\theta=0 \quad \text{on } L^{(2)}(U),$$
$$\bar\partial_\tau\Psi_\pm+[\theta,\Psi_\pm]=0,\quad \bar\partial_\tau X_\pm+[\theta,X_\pm]=0 \quad \text{on } L^{(1)}(U), \tag{2.10}$$
$$\bar\partial_\tau\Phi_\pm+[\theta,\Phi_\pm]=\{\Psi_\pm,X_\pm\},$$
$$\bar\partial_\tau\Omega+[\theta,\Omega]=2\{\Psi_+,X_-\}+2\{X_+,\Psi_-\} \quad \text{on } L(U)$$

are satisfied and, secondly, we have

$$\mathscr{P}\theta=a,\quad \mathscr{P}\Psi_\pm=\psi_{A\pm},\quad \mathscr{P}X_\pm=\chi_{A\pm}, \tag{2.11}$$
$$\tilde{\mathscr{P}}\Phi_\pm=\varphi_\pm,\quad \tilde{\mathscr{P}}\Omega=[\varphi_+,\varphi_-],$$

where $\tilde{\mathscr{P}}$ is the suitable modification of $\mathscr{P}$, accounting non-closeness of forms $\Phi_\pm$ and Ω.

On the basis of [15], [32] the following result was obtained in [18].

THEOREM 2.4a ([18]). *In order that the fields a, $\psi_{A\pm}$ $\chi_{A\pm}$, $\varphi_\pm$ on U satisfy the super-symmetric Yang–Mills equations* (2.9) *it is necessary and sufficient that the fields θ, $\Psi_\pm$, $X_\pm$, $\Phi_\pm$, Ω corresponding to them in virtue of the relations* (2.10), (2.11), *satisfy the Cauchy–Riemann equations of the form:*

$$\bar\partial\Phi_\pm+[\theta,\Phi_\pm]+\{\Psi_\pm,X_\pm\}-\tfrac12[\Omega,\Phi_\pm]\langle z\cdot w\rangle=0,$$
$$\bar\partial\Omega+[\theta,\Omega]+2\{\Psi_+,X_-\}+2\{X_+,\Psi_-\}+$$
$$+[\Phi_+,\Phi_-]\langle z\cdot w\rangle=0 \quad \text{on } L^{(1)}(U),$$
$$\left.\begin{aligned}\bar\partial\Psi_\pm+[\theta,\Psi_\pm]+\tfrac12[\Phi_\pm,\Psi_\mp]\langle z\cdot w\rangle=0,\\ \bar\partial X_\pm+[\theta,X_\pm]+\tfrac12[\Phi_\pm,X_\mp]\langle z\cdot w\rangle=0,\end{aligned}\right\} \tag{2.12}$$
$$\text{on } L^{(2)}(U),$$
$$\bar\partial\theta+\theta\wedge\theta+\tfrac16(\bar\partial\Omega+[\theta,\Omega])\langle z\cdot w\rangle+\tfrac13(\{\Psi_+,X_-\}+\{X_+,\Psi_-\})\langle z\cdot w\rangle+$$
$$+\tfrac16[\Phi_+,\Phi_-]\langle z\cdot w\rangle^2$$
$$\text{on } L^{(3)}(U).$$

At first sight the relations (2.12) seem to be as complicated as the equations (2.9). However, the equations (2.12), and not (2.9), are the ones which have a clear geometrical meaning.

In fact, following A. Ferber [10] and E. Witten [50], let us consider first the projective spaces of super-twistors $\mathscr{L}_+^N$ (resp. dual super-twistors $\mathscr{L}_-^N$) with four even coordinates z^A, $z_{B'}$ (resp. w_A, $w^{B'}$) and N odd coordinates ζ^k (resp. η^k), and secondly a super-manifold $\mathscr{L}_N(U)=\{(z,\zeta;$

$w, \eta) \in \mathscr{L}_+^N \times \mathscr{L}_-^N$: $\langle z \cdot w\rangle = \zeta^k \eta_k$, $(z, w) \in L(U)\}$, parametrizing the analogues of light rays on the Minkowski super-space.

We shall examine the CR-submanifolds of the form

$$L_+^N = \{(z, \zeta) \in \mathscr{L}_+^N: \operatorname{Im}(z^0 \bar{z}_{0'} + z^1 z_{1'}) = \zeta^k \bar{\zeta}_k\},$$

$$L_-^N = \{(w, \eta) \in \mathscr{L}_-^N: \operatorname{Im}(w_0 \bar{w}^{0'} + w_1 \bar{w}^{1'}) = \eta^k \bar{\eta}_k\},$$

$$L_N(U) = (L_+^N \times L_-^N) \cap \mathscr{L}_N(U)$$

on these super-manifolds.

We shall denote by $L_N^{(j)}(U)$ the j-th infinitesimal neighbourhood of the manifold $L_N(U) \subset L_+^N \times L_-^N$. Now, we shall examine on the CR-manifold $L_+^2 \times L_-^2$ the $(0, 1)$-form Θ given by

$$\begin{aligned} \Theta = \theta + \Psi_+ \zeta_1 + X_+ \zeta_2 + \Psi_- \eta_2 + X_- \eta_1 + \\ + \Phi_+ \zeta_1 \zeta_2 + \Phi_- \eta_1 \eta_2 + \Omega(\zeta_1 \eta_1 + \zeta_2 \eta_2), \end{aligned} \tag{2.13}$$

where the forms $\theta, \Psi_\pm, X_\pm, \Phi_\pm$ and Ω satisfy the relations (2.10). In virtue of (2.10) we have $\bar{\partial}_\tau \Theta + \Theta \wedge \Theta = 0$ on $L(U)$, i.e., the form Θ defines a CR-fibre bundle over $L_2(U)$ which is trivial over any quadric $L(u)$, $u \in U$.

THEOREM 2.4b ([18]). *In order that the form Θ given by* (2.13) *be gauge-equivalent to a form satisfying the equation*

$$\bar{\partial}_\tau \Theta + \Theta \wedge \Theta = O\big((\langle z \cdot w\rangle - \zeta^k \eta_k)^2\big) \tag{2.14}$$

it is necessary and sufficient that the components $(\theta, \Psi_\pm, X_\pm, \Phi_\pm, \Omega)$ of the form Θ satisfy the Cauchy–Riemann equations of the form (2.12).

The equality (2.14) means that the form Θ defines a CR-fibre bundle over $L_2^{(1)}(U)$.

From Theorems 2.4a and 2.4b and their analogues for other super-symmetrical Yang–Mills theories results the following

THEOREM 2.4 (E. Witten [50], G. M. Henkin [18]). *The Penrose transformation establishes an isomorphism of the space of CR-fibre bundles over $L_N^{(3-N)}$, $N = 1, 2, 3$, trivial on all quadrics $L(u)$, $u \in U$, with the space of (smooth) solutions of the N-symmetrical system of Yang–Mills equations.*

For $N = 3$ (and for holomorphic fields and fibre bundles) this result was obtained in a paper by E. Witten [50]. More exactly, in [50] the equations of motion of the 3-supersymmetrical Yang–Mills system were reinterpreted as the conditions for the integrability of a connection along super-light rays in the Minkowski super-space. An extension of Witten's result to the case of the remaining $N = 1, 2, 4$ was obtained in [18].

For the formulation of this result in the case of $N = 4$ it is necessary to employ a natural extension of the well-known twistor transformation (see [8]) $\mathscr{T}$ establishing a canonical isomorphism between the spaces $H^1(L_+(U),\ \mathcal{O}(2s-2,0))$ and $H^1(L_-(U),\ \mathcal{O}(0,\ -2s-2))$, $s = 0, \frac{1}{2}, 1$, onto the CR-fibre bundles over $L_4(U)$.

THEOREM 2.5 ([18]). *The Penrose transformation establishes an isomorphism between the space of CR-fibre bundles over $L_4(U)$ invariant with respect to the twistor transformation $\mathscr{T}$ and trivial on all quadrics $L(u)$, $u \in U$, and the space of all (smooth) solutions on U of the 4-supersymmetrical Yang–Mills system.*

For $N = 3, 4$, recently, A. A. Rosly [39] has discovered another interpretation of the Yang–Mills relations as integrability conditions along some tangent subspaces (of purely odd dimension) of the Minkowski super-space.

In comparison with Theorem 2.4, Theorems 2.4a and 2.4b from [18] contain additional information — a twistor interpretation of all fields and equations entering into the super-symmetrical Yang–Mills system. Let us note that these results give development of a paper by Yu. I. Manin [31], where the cohomological component analysis is given on $\mathscr{L}_3(U)$ of the 3-super-symmetrical Yang–Mills equations on U.

Interpretations in terms of Cauchy–Riemann equations over twistors of some other classical (not super-symmetrical) interactions between the Yang–Mills–Higgs and Dirac fields are obtained in papers [15], [18], [29].

We have touched here only part of the works dealing with a twistor interpretation of gauge fields on a plane Minkowski space. The problem of establishing the twistor theory for non-plane Minkowski spaces was discussed in a very impressive paper by R. Penrose [36]. Further developments of this work were obtained by C. Le Brun [27], [28], T. Isenberg, Ph. Yasskin [24], Yu. I. Manin, I. Penkov [33].

The twistor theory yields, as we know great results in establishing exact (and physically interesting) solutions of the self-dual Einstein and the Yang–Mills equations (see M. F. Atiyah [3], R. Penrose [36], R. Ward [48] and others). One can hope that the twistor interpretation of non-self-dual equations can also lead, to some new interesting solutions. The first non-trivial investigations of this kind are contained in papers by P. Forgacs, Z. Horvath, L. Palla [11] and Yu. I. Manin [31].

References

[1] Andreotti A. and Hill C. D., *Ann. Scuola Norm. Sup. Pisa* **26** (1972), pp. 325–363 and 747–806.
[2] Andreotti A., Fredricks G. and Nacinovich M., *Ann. Scuola Norm. Sup. Pisa, ser. IV*, **8** (3) (1981), pp. 365–404.
[3] Atiyah M. F., *Geometry of Yang–Mills Fields*, Lexiconi Fermianici, Pisa, 1979.
[4] Ayrapetian R. A. and Henkin G. M., *Doklady AN SSSR* **259** (4) (1981), pp. 777–781; *Uspekhi Matem. Nauk* **39** (3) (1984), pp. 39–106.
[5] Bedford E., *Math. Ann.* **230** (1977), pp. 213–225.
[6] Bogges A., *Duke Math. J.* **47** (4) (1980), pp. 903–921.
[7] Eastwood M. G., *Proc. Roy. Soc. London* **A 374** (1981), p. 431.
[8] Eastwood M. G. and Ginsberg M. C., *Duke Math. J.* **48** (1981), p. 177.
[9] Eastwood M. G., Penrose R. and Wells R. O., *Commun. Math. Phys.* **78** (1981), pp. 305–351.
[10] Ferber A.. *Nuclear Physics*, **B 132** (1978), pp. 55–64.
[11] Forgacs P., Horvath Z., and Palla L., *Physics Letters* **115B** (6) (1982), pp. 463–467.
[12] Gindinkin S. G. and Henkin G. M., In: *Sovremennye problemy matematiki* vol. 17, WINITI, Moscow, 1981, pp. 57–112.
[13] Henkin G. M.. *Uspekhi Matem. Nauk* **32** (3) (1977), pp. 57–118.
[14] Henkin G. M., *Doklady AN SSSR* **255** (4) (1980), pp. 844–847.
[15] Henkin G. M., *Doklady AN SSSR* **260** (5) (1981), pp. 1086–1089.
[16] Henkin G. M., *C. R. Acad. Sc., Paris* **292** (1981), pp. 27–30.
[17] Henkin G. M., *Mathematical Review* **82E** (1982), pp. 2222–2224.
[18] Henkin G. M., *Trudy II Mezhdunarodnogo Seminara "Teoretikogruppovye metody v fizike", Zvenigorod, 24–26 noyabr' 1982*, Nauka, Moscow 1983, pp. 107–116.
[19] Henkin G. M., *Doklady AN SSSR* **274** (3) (1984), pp. 553–558.
[20] Henkin G. M. and Manin Yu. I., *Physics Letters* **95B** (1980), pp. 405–408.
[21] Henkin G. M. and Manin Yu. I., *Comp. Math.* **44** (1981), pp. 103–111.
[22] Hitchin N. T., *Proc. Roy. Soc. London* **A 370** (1980), pp. 173–183.
[23] Isenberg T. and Yasskin Ph., *Gen. Rel. Grav.* **14** (1982), pp. 621–627.
[24] Isenberg T., Yasskin Ph. and Green P., *Physics Letters* **78B** (1978), pp. 462–467.
[25] Kohn J. J., *Proc. Symp. Pure Math.* **30** (1977), pp. 215–237.
[26] Kohn J. J. and Rossi H., *Ann. of Math.* **81** (1965), pp. 451–472.
[27] Le Brun C., *Letters in Math. Phys.* **6** (1982), pp. 345–354.
[28] Le Brun C., *Proc. R. Soc. London* **A 380** (1982), pp. 171–175.
[29] Leiterer J., *Math. Nachr.* **112** (1983), pp. 35–67.
[30] Manin Yu. I., In: *Sovremennye problemy matematiki*, vol. 17, WINITI, Moscow, 1981, pp. 3–56.
[31] Manin Yu. I., *Trudy MIAN* **165** (1984), pp. 98–114.
[32] Manin Yu. I. and Henkin G. M., *Yadernaya fizika* **35** (6) (1982), pp. 1610–1626.
[33] Manin Yu. I. and Penkov I. B., *Funktsionalnyi analiz i ego prilozheniya* **16** (1) (1982), pp. 78–79.
[34] Naruki I., *Publ. Res. Inst. Math. Kyoto Univ.* **8** (1) (1972), pp. 43–110.
[35] Naruki I., *Publ. Res. Inst. Math., Kyoto Univ.* **6** (1970), pp. 113–187.
[36] Penrose R., *Gen. Rel. Grav.* **7** (1976), pp. 31–52 and 171–176.
[37] Penrose R., *Reports on Math. Phys.* **12** (1977), pp. 65–76.

[38] Penrose R., *Research Notes in Mathematics* **37**, Pitman Advanced Publishing Psogram, 1979, p. 88–93.
[39] Roslyi A. A., *Trudy II Mezhdunarodnogo Seminara "Teoretikogruppovye metody v fizike", Zvenigorod, 24–26, noyabr' 1982,* Nauka, Moscow, 1983, pp. 263–268.
[40] Rossi H. and Vergne M., *Pacif. J. of Math.* **65** (1) (1976), pp. 193–207.
[41] Sato M., Kawai T. and Kashivara M., *Lecture Notes in Mathematics* **287**, Springer-Verlag, 1973, pp. 265–529.
[42] Sibony N., *Lecture Notes in Mathematics* **578**, Springer-Verlag, 1977, pp. 14–27.
[43] Spencer D. C., *Bull. Amer. Math. Soc.* **75** (1969), pp. 179–239.
[44] Treves F., *Ann. of Math.* **104** (1976), pp. 269–324.
[45] Treves F., Éc. Polytechnique Centre Math. publications, Mai, 1981.
[46] Vladimirov W. S., *Lecture Notes in Mathematics* **919**, Springer-Verlag, 1982.
[47] Ward R., *Physics Letters* **61A** (1977), p. 81.
[48] Ward R., *Commun. Math. Phys.* **80** (1981), pp. 563–574.
[49] Wells R. O., *Commun. Math. Phys.* **78** (1981), pp. 567–600.
[50] Witten E., *Physics Letters* **77B** (1978), pp. 394.

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

PETER W. JONES

Recent Advances in the Theory of Hardy Spaces

This paper is a survey of recent results in the theory of Hardy spaces and BMO. We denote by H^p the Hardy space of analytic functions on $\boldsymbol{R}_+^2 = \{z = x+iy: y > 0\}$, and we denote by $\mathscr{H}^p(X)$ the real variables Hardy space of functions on X. Results related to those discussed in this paper can be found in the papers of Jean Bourgain and Yves Meyer.

1. Interpolating sequences and the $\bar\partial$ problem on $\boldsymbol{R}_+^2$

A sequence $\{z_j\}$ of points in $\boldsymbol{R}_+^2$ is called an *interpolating sequence* if whenever $\{a_j\} \in l^\infty$ there is $F \in H^\infty$ such that $F(z_j) = a_j$. By a theorem due to Carleson [4], $\{z_j\}$ is an interpolating sequence if and only if the points are uniformly separated in the hyperbolic metric and the measure $\sum\limits_j y_j \delta_{z_j}$ is a Carleson measure. (μ is a Carleson measure if $\|\mu\|_C \equiv \sup\limits_I |I|^{-1} |\mu|(\{I \times \times (0, |I|]\}) < \infty$, the sup taken over all intervals $I \subset \boldsymbol{R}$.) If $\{z_j\}$ is an interpolating sequence, P. Beurling's theorem (see [5]) asserts that there are functions $F_j \in H^\infty$ such that $F_j(z_k) = \delta_{j,k}$ and $\sup\limits_z \sum\limits_j |F_j(z)| < \infty$. Thus interpolation may be performed linearly by mapping $\{a_j\}$ to $\sum\limits_j a_j F_j$. Interpolating sequences are intimately related to solutions of the $\bar\partial$ problem. A fundamental link is provided by the following lemma of [20].

LEMMA 1. *Every Carleson measure can be represented as a (constant multiple of a) weak star limit of finite convex combinations of measures of the form $\sum a_j y_j \delta z_j$, where $\|\{a_j\}\|_{l^\infty} \leqslant 1$ and $\{z_j\}$ is an interpolating sequence. (The interpolating constants associated to the sequences $\{z_j\}$ are uniformly bounded.)*

Suppose we wish to solve $\bar\partial F = \mu$ with $F \in L^\infty$ on $\boldsymbol{R}$ and suppose μ is of the form $\sum a_j y_j \delta_{z_j}$ where $\{a_j\} \in l^\infty$ and $\{z_j\}$ is an interpolating sequence.

Let B be the Blaschke product with simple zeros at the points $\{z_j\}$. We can then find an H^∞ function G which interpolates the correct values on $\{z_j\}$ so that $\bar\partial(G/B) = \mu$. On $\boldsymbol{R}$ we have $\|G/B\|_{L^\infty(\boldsymbol{R})} = \|G\|_{H^\infty} \leqslant C\|\{a_j\}\|_{l^\infty}$. (This line of reasoning is just as easily reversed to obtain H^∞ interpolating functions from solutions of the $\bar\partial$ problem.) If we wish to solve the general problem $\bar\partial F = \mu$ with $F \in L^\infty(\boldsymbol{R})$ and where μ is a Carleson measure we can simply combine the above reasoning with Lemma 1. This reasoning was first used in [20] to give a constructive solution of the $\bar\partial$ problem in $\boldsymbol{R}^2_+$ which yields solutions which are bounded on $\boldsymbol{R}$ when the data is a Carleson measure. To carry this out one must also have a constructive way of building H^∞ interpolating functions; such a method was first provided by Earl [9].

A more efficient way to solve interpolation or $\bar\partial$ problems is to have closed formulae. We first show how to obtain P. Beurling functions for an interpolating sequence $\{z_j\}$. (See [22] for details.) Let B_j be the Blaschke product with simple zeros at $\{z_k\colon k \neq j\}$ and let $\varepsilon = \inf_j |B_j(z_j)|$. (Since $\{z_j\}$ is an interpolating sequence, $\varepsilon > 0$.) Now set

$$F_j(z) = c_j B_j(z)\left(\frac{y_j}{z-\bar z_j}\right)^2 \exp\left\{\frac{-i}{\log 2/\varepsilon}\sum_{y_k\leqslant y_j}\frac{y_k}{z-\bar z_k}\right\},$$

where c_j is chosen so that $F_j(z_k) = \delta_{j,k}$. Standard arguments show that $|c_j| \leqslant C(\varepsilon)$. To show that $\sup_z \sum_j |F_j(z)| \leqslant C'(\varepsilon)$ it is sufficient by the maximum principle to take $z = x \in \boldsymbol{R}$ and (since $\|B_j\|_{H^\infty} = 1$) evaluate

$$\sum_j \left|\frac{y_j}{x-\bar z_j}\right|^2 \exp\left\{\operatorname{Re}\left(\frac{-i}{\log 2/\varepsilon}\sum_{y_k\leqslant y_j}\frac{y_k}{x-\bar z_k}\right)\right\}$$
$$= \sum_j \left|\frac{y_j}{x-\bar z_j}\right|^2 \exp\left\{\frac{-1}{\log 2/\varepsilon}\sum_{y_k\leqslant y_j}\left|\frac{y_k}{x-\bar z_k}\right|^2\right\}.$$

The last sum is easily seen to be a lower Riemann sum for $\int_0^\infty \exp\left\{\frac{-t}{\log 2/\varepsilon}\right\}dt = (\log 2/\varepsilon)^{-1}$. (The constant $\log 2/\varepsilon$ is inserted so as to make the F of minimal norm.)

Much the same procedure can be used to find solutions to the $\bar\partial$ problem. For a positive measure σ on $\boldsymbol{R}^2_+$ set

$$K(\sigma, z, \zeta) = \frac{2i}{\pi(z-\zeta)}\left(\frac{\operatorname{Im}\zeta}{z-\bar\zeta}\right)\exp\left\{\int_{\{\operatorname{Im}(w)\leqslant\operatorname{Im}(\zeta)\}}\left(\frac{-i}{z-\bar w}+\frac{i}{\zeta-\bar w}\right)d\sigma(w)\right\}.$$

The following result is also contained in [22]; its proof is virtually identical to the argument given in the last paragraph.

THEOREM 1. *If μ is a Carleson measure and*

$$F(z) = \int_{\mathbf{R}^2_+} K(|\mu|/\|\mu\|_C, z, \zeta)\, d\mu(\zeta),$$

$F(z) \in L^1$ loc. *on* $\mathbf{R}^2_+$ *and* $\bar{\partial} F = \mu$. *Furthermore, whenever* $x \in \mathbf{R}$,

$$\int_{\mathbf{R}^2_+} |K(|\mu|/\|\mu\|, x, \zeta)|\; d|u|(\zeta) \leqslant C\|\mu\|_C.$$

It should be pointed out that no analogue of Theorem 1 can exist for the ball in $\mathbf{C}^n$, $n > 1$. Varopoulos [34] has constructed (in dimension greater than one) a $\bar{\partial}$ closed form which satisfies a Carleson condition but admits no $\bar{\partial}$ solution which is bounded on the boundary.

2. Applications of solutions to $\bar{\partial}$

On of the more interesting problems considered in the past few years is how to obtain constructively the Fefferman–Stein decomposition of a BMO function. In the one dimensional version this means taking $\varphi \in$ BMO and constructing $u, v \in L^\infty$ such that $\varphi = u + Hv$ ($+$ constant), where H is the Hilbert transform. To this end fix $\varphi \in$ BMO, φ real valued, and use Varopoulos' (constructive) method [34] to extend φ to $\Phi \in C^\infty(\mathbf{R}^2_+)$ so that $|\bar{\partial}\Phi|\; dx\, dy$ is a Carleson measure. Suppose that F satisfies $\bar{\partial} F = \bar{\partial}\Phi$ and F is bounded on $\mathbf{R}$. Then since $G \equiv \Phi - F$ is holomorphic the boundary values satisfy $H(\operatorname{Re} G) = \operatorname{Im} G$, or equivalently $\varphi = \operatorname{Re} F + H(\operatorname{Im} F)$. This line of reasoning was first used in [20] to find F and obtain the BMO splitting. Setting $\mu = \bar{\partial}\Phi\, dx\, dy$ and invoking Theorem 1 provides another method for finding F. (See [22] for details.)

Theorem 1 is useful in practice because information on the size and smoothness of F can easily be obtained from similar information on μ. It can be used for instance to decompose $f \in H^p$ as a sum of $f_1 \in H^{p/2}$ and $f_2 \in H^\infty$. (See Theorem 2 of [22] for a precise statement.) We give two examples related to this phenomenon. Let $(\cdot, \cdot)_{\theta,q}$ and $(\cdot, \cdot)_\theta$ denote respectively the real and complex method of interpolation of (quasi) normed vector spaces. The following two results are contained in [22]. See [23] for a survey of interpolation between Hardy spaces. We denote by $H^{p,q}$ the space of all holomorphic functions whose non-tangential maximal functions lie in the Lorentz space $L^{p,q}$. (So $H^{p,p} = H^p$.)

THEOREM 2. $(H^{p_0}, H^\infty)_{\theta,q} = H^{p,q}$, $\dfrac{1}{p} = \dfrac{1-\theta}{p_0}$,
$0 < p_0 < \infty$, $0 < \theta < 1$, $0 < q \leqslant \infty$.

THEOREM 3. $(H^{p_0}, H^\infty)_\theta = H^p$, $\dfrac{1}{p} = \dfrac{1-\theta}{p_0}$,
$0 < p_0 < \infty$, $0 < \theta < 1$.

Since Hörmander's paper [17] it has been well known that L^∞ solutions of the $\bar\partial$ problem are closely related to the corona problem. In 1980 T. Wolff (see [14]) used the following theorem to give an elementary proof of Carleson's corona theorem for the disk.

THEOREM 4. *Suppose* $\| \, |U|^2 y \, dx\, dy \|_C \leqslant Z A_1$ *and* $\| \, |\partial U| \, y \, dx \, dy \, \|_C \leqslant A_2$. *Then there is* F *such that* $\bar\partial F = U$ *and* $\|F(x, 0)\|_{L^\infty(\mathbf{R})} \leqslant C(A_1, A_2)$.

It is an unfortunate state of affairs that the only known proofs of Theorem 4 use either duality or the full Uchiyama machine (see the next section). It would be very helpful in the study of H^∞ on planar domains or more general Riemann surfaces to have a formula in the spirit of Theorem 1 which could be used to solve Theorem 4. On the positive side, the idea behind Theorem 1 can be used to prove the corona theorem for certain classes of planar domains. We say that a domain $\mathscr{D} = \mathbf{C} \setminus E$, where $E \subset \mathbf{R}$, is *homogeneous* if there is $\varepsilon > 0$ such that for all $x \in E$ and $\delta > 0$, $|E \cap (x-\delta, x+\delta)| \geqslant \varepsilon\delta$. Carleson [7] has shown that whenever $\mathscr{D}$ is homogeneous the corona theorem holds. In fact Carleson proves much more. Let Δ, the unit disk, be the covering surface of $\mathscr{D}$, and let Γ be the group of deck transformations. By $H^\infty(\Gamma)$ we denote the intersection of $H^\infty(\Delta)$ and Γ invariant functions, so that $H^\infty(\mathscr{D}) \cong H^\infty(\Gamma)$. A bounded linear operator P: $H^\infty(\Delta) \to H^\infty(\Gamma)$ is called a Γ *projection operator* if $P^2 = P$, $P(1) = 1$, and $P(fg) = fP(g)$ for $f \in H^\infty(\Gamma)$, $g \in H^\infty(\Delta)$. Such projection operators were first considered by Forelli [12]. Carleson proved a concrete version of Forelli's theorem by showing that when $\mathscr{D}$ is a homogeneous domain, a Γ projection operator exists. The proof consists of first setting up an appropriate $\bar\partial$ problem. One then uses formulae like those in Theorem 1 in combination with some difficult estimates on harmonic measure to solve that $\bar\partial$ problem with good bounds.

Recently, the authors of [25] have been able to simplify and slightly generalize Carleson's results by using the projection operator constructed by Forelli [12] (see also [10]). Let B be the Blaschke product $\prod_{\gamma\in\Gamma} \gamma(z)\bar{\gamma}(0)/|\gamma(0)|$ associated to a Fuchsian group Γ and let $\{z_j\} = \{z\colon B'(z) = 0\}$.

LEMMA 2. *Suppose Green's function tends to zero at the boundary of Δ/Γ. If $\{z_j\}$ is an interpolating sequence, there is a Γ projection operator.*

A corollary of Lemma 2 is the corona theorem for $H^\infty(\Gamma)$. (When Green's function tends to zero at the boundary it is shown that the corona theorem is equivalent to the following condition: whenever $f_1, \ldots, f_N \in H^\infty(\Gamma)$ are corona data, there are $g_1, \ldots, g_N \in H^\infty(\Delta)$ such that $\sum_k f_k g_k \equiv 1$ and $g_k(\gamma(z_j)) = g_k(z_j)$ for all j, k, and $\gamma \in \Gamma$.) To obtain Carleson's results on homogeneous domains then becomes a matter of establishing some elementary estimates on harmonic measure to show that Lemma 2 can be invoked.

An alternative method for solving the corona problem on homogeneous domains is to solve the $\bar\partial$ problem $\bar\partial F = b\, dx$ on $\mathscr{D}$, where $b \in L^\infty$ is supported on $\boldsymbol{R}\setminus E$ and where we demand that F be bounded on E. This can be done directly (see [25]) by combining the formula from Theorem 1 with some estimates on harmonic measure and a result due to Pommerenke [28]. Pommerenke's result asserts that a Riemann surface Ω is of Widom type (see [35]) if and only if the Blaschke product B associated to Γ is such that B' is in the Nevanlinna class. In that case, Pommerenke sets $G(z)$ as the inner factor of B' and defines an operator

$$Tf(z) = \frac{B(z)}{B'(z)} \sum_{\gamma\in\Gamma} G(\gamma(z))\, f(\gamma(z))\, \frac{\gamma'(z)}{\gamma(z)}$$

which maps $H^\infty(\Delta) \to H^\infty(\Gamma)$. This operator satisfies $Tf(0) = f(0)$ and $Tf(z) \leqslant \sup_\gamma |f(\gamma(z))|$ and thus provides an excellent method for building auxiliary functions in $H^\infty(\Omega)$. (Unfortunately $T(1) \neq 1$.) It would be interesting to have a better understanding of the relation between solutions of the $\bar\partial$ problem and Pommerenke's operators in light of the following fact: B. Cole (see [13]) has constructed a Riemann surface for which the corona theorem fails. It is not hard to see that Cole's example can be modified so as to be of Widom type.

3. BMO

We return to the Fefferman–Stein decomposition of BMO functions. We have already seen that, in dimension one, complex function theory is a useful tool in obtaining the F.–S. decomposition. Another example of this is furnished by the argument of [8]. Let $\varphi \in \mathrm{BMO}$ be real valued and of small enough norm so that $w = e^{\varphi}$ lies in the Muckenhoupt A_2 class. Define an operator S by $S(f) = w^{-1/2}|H(fw^{1/2})| + w^{1/2}|H(fw^{-1/2})|$. By the Hunt–Muckenhoupt–Wheeden theorem [18] the function U $U = \sum_{j=0}^{\infty} C^{-j}\cdot S^{j}(\chi_{[0,1]})$ lies in $L^2(dx)$ as soon as C is large enough. Let $w_1 = Uw^{1/2}$, $w_2 = Uw^{-1/2}$. and set $W_j = w_j + iH(w_j)$. Then since $S(U) \leqslant CU$ pointwise,

$$\varphi = \log w = \left(\log w - \log\left|\frac{W_1}{W_2}\right|\right) + H\left(-\arg\left(\frac{W_1}{W_2}\right)\right) = u + Hv,$$

where $\|u\|_{L^\infty} \leqslant 2\log(1+C)$ and $\|v\|_{L^\infty} \leqslant 2\arctan(C) < \pi$. (This procedure thus provides a passage from the H. M. W. theorem to a weak form of the Helson–Szegö theorem [16].) It is unfortunate that it is difficult to deduce any information on u and v from given properties of φ when using this method. The iterative procedure used to produce the above function U is based on an idea due to Rubio de Francia [29] which can be used ([29], [8]) to give a short proof of the factorization theorem [21] for Muckenhoupt A_p weights:

$$w \in A_p \text{ if and only if } w = w_1(w_2)^{1-p}, \quad w_j \in A_1.$$

The problem with the preceding methods of decomposing BMO functions lies in the fact that those methods cannot be generalized to higher dimensions. In an extremely important paper [31], Uchiyama presents a new constructive machine for decomposing BMO functions. By C. Fefferman's duality theorem [11], every function $\varphi \in \mathrm{BMO}(\boldsymbol{R}^n)$ is of the form $u_0 + \sum_{j=1}^{n} R_j u_j$ (+ constant) where $u_0, \ldots, u_n \in L^\infty$ and the R_j's are the Riesz transforms. Thus the Riesz transforms characterize $\mathrm{BMO}(\boldsymbol{R}^n)$ (equivalently $\mathscr{H}^1(\boldsymbol{R}^n)$), but it is not at all clear from Fefferman's (duality) proof which families of Calderón–Zygmund kernels can replace the Riesz kernels. The answer is provided by Uchiyama's

THEOREM 5. *Suppose $\theta_1, \ldots, \theta_m$ are $C^\infty(\boldsymbol{R}^n\setminus\{0\})$ Fourier multipliers which are homogeneous of degree zero. Then every $\varphi \in \mathrm{BMO}(\boldsymbol{R}^n)$ is of the form $\varphi = \sum_{j=1}^{m} (\theta_j\cdot\hat{u}_j)^{\vee}$, $u_1, \ldots, u_m \in L^\infty$, if and only if*

$$\operatorname{Rank}\begin{bmatrix}\theta_1(x) & \ldots & \theta_m(x)\\ \theta_1(-x) & \ldots & \theta_m(-x)\end{bmatrix} = 2, \quad x \in \boldsymbol{R}^n\setminus\{0\}. \tag{U}$$

The assertion that (U) must hold if the decomposition always exists is due to Janson [19]. In case (U) holds, Uchiyama shows how to constructively decompose a given $\varphi \in \mathrm{BMO}$. One starts by taking a C^∞ function ψ with support in $\{|x| \leqslant 1\}$ and satisfying $\int_0^\infty \hat{\psi}(\xi t)^2 t^{-1} dt = 1$ for all $\xi \in \boldsymbol{R}^n \setminus \{0\}$, and then using the identity

$$\varphi(x) = \int_0^\infty (\psi_t * \psi_t * \varphi)(x) t^{-1} dt = \sum_{N=-\infty}^{\infty} \int_{2^N}^{2^{N+1}} \equiv \sum_{N=-\infty}^{\infty} \varphi_N(x).$$

It is sufficient to decompose $\sum_{N=-M}^{\infty} \varphi_N$, for then one can just take limits. Uchiyama now builds by induction a vector valued function $\beta_k = (\beta_{k,1}, \ldots$ $\ldots, \beta_{k,m})$ such that $\sum_{j=1}^{m} (\theta_j \cdot \hat{\beta}_{k,j})^\vee \approx \sum_{N=-M}^{k} \varphi_n$, such that $\sum_{j=1}^{m} |\beta_{k,j}(x)|^2 \equiv R^2$ where R is some large fixed constant, and where β_k has some smoothness properties. The central idea of Uchiyama is that one can construct functions $a_{k+1,1}, \ldots, a_{k+1,m}$, each essentially of the same type as the function φ_{k+1}, such that $\varphi_{k+1} \approx \sum_{j=1}^{m} (\theta_j \cdot \hat{a}_{k+1,j})^\vee$, and such that

$$\beta_k(x) \cdot \big(a_{k+1,1}(x), \ldots, a_{k+1,m}(x)\big) = \beta_k(x) \cdot a_{k+1}(x) \approx 0$$

whenever $x \in (2^{-k-1} \cdot \boldsymbol{Z})^n$. It is for this last orthogonality condition that condition (U) is used in a crucial manner. Uchiyama now sets $\beta^{k+1} = R \cdot \dfrac{\beta_k + a_{k+1}}{|\beta_k + a_{k+1}|}$. The orthogonality condition between β_k and a_{k+1}, and the smoothness condition on β_k can now be used to deduce that $\beta_{k+1} - (\beta_k + a_{k+1})$ vanishes more like $R^{-1}(a_{k+1})^2$ than a_{k+1}. This last fact allows one to show that

$$\Big\| \sum_{N=-M}^{\infty} \varphi_N - \sum_{j=1}^{m} (\theta_j \cdot \hat{\beta}_{\infty,j})^\vee \Big\|_{\mathrm{BMO}} \leqslant \text{const } R^{-1} \|\varphi\|_{\mathrm{BMO}},$$

where $\beta_{\infty,j} = \lim_k \beta_{k,j}$. An iteration argument now completes the proof of the theorem. The proof outlined above uses some ideas developed by Uchiyama [33] to constructively split BMO functions in the (easier) martingale case.

Uchiyama has recently extended these results in [33] where he proves the following singular integral characterization of H^p. We denote by $F(x, y)$ the Poisson extension to $\boldsymbol{R}_+^{n+1} = \{(x, y)\colon x \in \boldsymbol{R}^n,\ y > 0\}$ of $F(x)$.

THEOREM 6. *If* (U) *holds there is* $p_0 = p_0(\theta_1, \ldots, \theta_m) < 1$ *such that whenever* $p_0 < p \leqslant 1$,

$$\|f\|_{\mathscr{H}^p(\mathbf{R}^n)} \sim \sup_{y>0} \sum_{j=1}^{m} \|(\theta_j \cdot \hat{f})^{\vee}(x, y)\|_{L^p(dx)}.$$

Theorem 6 has been known for the Riesz system since the 1960 paper of Stein and Weiss [30]. Stein and Weiss used subharmonicity properties of the Riesz system in their proof. Uchiyama proves a real variables analogue of subharmonicity to obtain his result. This analogue of subharmonicity is an easy consequence of a difficult technical theorem on the decomposition of certain weighted BMO spaces. This last result turns out to be a hybrid of Theorems 1 and 5; see [32] for a full discussion. It would be very interesting to see if Uchiyama's techniques could be altered to show (or disprove) that when condition (U) holds, $\mathscr{H}^p$ can be characterized by the corresponding singular integrals for *all* $p > 0$. This problem is open even for the Riesz system. (See however [3] for some related results.)

4. Banach space properties

Let $\mathscr{H}^p_d$ denote the dyadic (martingale) Hardy space of functions on $[0, 1]$. B. Maurey proved in [27] that $\mathscr{H}^1_d$ is isomorphic as a Banach space with all of the spaces $\mathscr{H}^1(\mathbf{R}^n)$, $n \geqslant 1$. (Equivalently, all of these BMO spaces are isomorphic.) Maurey's proof heavily used Banach space theory. Carleson [6] was the first to exhibit an explicit unconditional basis for $\mathscr{H}^1(\mathbf{R}^n)$; the existence of this basis is of course equivalent to Maurey's theorem. Wojtaszczyk [36] later showed that the Franklin functions also provide an unconditional basis. Wojtaszczyk's proof could be seen as an extension of Bočkarev's method [2] of showing that the (holomorphic) Franklin functions are a basis for the disk algebra $= H^\infty \cap$ continuous functions.

In [24] it is shown that BMO (equivalently L^∞/H^∞) has the bounded approximation problem. To do this it is easiest (though not necessary) to work with dyadic BMO $-$ BMO_d. The method of proof consists of taking $\varphi_1, \ldots, \varphi_M \in \mathrm{BMO}_d$ and building from them certain functions $\psi_1, \ldots, \psi_N$. Using standard Hilbert space notation, one then sets $Tf = \sum_{k=1}^{N} \langle f, \psi_k \rangle \|\psi_k\|_{L^2}^{-2} \cdot \psi_k$. If the ψ_k are chosen correctly, then N depends on M, $\|T\| \leqslant 20$, and $\|\varphi_j - T\varphi_j\| < \varepsilon$ for all j. (If ε is small enough, one can perturb the operator T to obtain the result.) The proof given in [24] uses some rather involved combinatorial arguments. It would be interesting to find a simpler proof which avoided this. Much more interesting would be a solution of the approximation problem for H^∞, the last Hardy space whose Banach space structure remains "not understood".

Let $H^p(B_n)$ denote the Hardy space of holomorphic functions on the unit ball in C^n. Wojtaszczyk [37] has shown that $H^1(B_1)$ is isomorphic as a Banach space to $H^1(B_n)$ for all n. An interesting aspect of his proof is that it uses inner functions on the ball, though this could be avoided. The existence of inner functions was proved by Aleksandrov [1] and, at about the same time, Løw [26]. Løw's paper was a modification of a construction due to Hakim and Sibony [15]. It seems clear that inner functions will play an important rôle in the study of the spaces $H^p(B_n)$.

References

[1] Alexandrov A. B., Existence of Inner Functions in the Ball (in Russian), *Mat. Sbornik* **118** (1982), pp. 147–163.

[2] Bočkarev S. V., Existence of a Basis in the Space of Functions Analytic in the Disk, and Some Properties of Franklin's System, *Mat. USSR Sbornik* **24** (1974), pp. 1–16.

[3] Calderón A. P. and Zygmund A., On Higher Gradients of Harmonic Functions, *Studia Math.* **24** (1964), pp. 211–226.

[4] Carleson L., An Interpolation Problem for Bounded Analytic Functions, *Amer. J. Math.* **80** (1958), pp. 921–930.

[5] Carleson L., Interpolation by Bounded Analytic Functions and the Corona Problem. In: *Proc. Int. Congr. Math.* (Stockholm, 1962), pp. 314–316.

[6] Carleson L., An Explicit Unconditional Basis in H^1, *Bull. des Sci. Math.* **104** (1980), pp. 405–416.

[7] Carleson L., On H^∞ in Multiply Connected Domains, In: *Conference on harmonic analysis in honor of Antoni Zygmund*, Vol. 2, Wadsworth Inc., Belmont, California, 1983, pp. 349–372.

[8] Coifman P. R., Jones P. W., and Rubio de Francia J. L., On a Constructive Decomposition of BMO Functions and Factorization of A_p Weights, *Trans. Amer. Math. Soc.*, to appear.

[9] Earl J. P., On the Interpolation of Bounded Sequences by Bounded Functions, *J. London Math. Soc.* **2** (1970), pp. 544–548.

[10] Earle C. J. and Marden A., On Poincaré Series With Applications to H^p Spaces on Bordered Riemann Surfaces, *Illinois J. Math.* **13** (1969), pp. 202–219.

[11] Fefferman C. and Stein E. M., H^p Spaces of Several Variables, *Acta Math.* **129** (1972), pp. 137–193.

[12] Forelli F., Bounded Holomorphic Functions and Projections, *Ill. J. Math.* **10** (1966), pp. 367–380.

[13] Gamelin T. W., *Uniform Algebras and Jensen Measures*, London Mathematical Society Lecture Note Series No. 32, 1978.

[14] Garnett J. B., *Bounded Analytic Functions*, Academic Press, New York, 1981.

[15] Hakim M. and Sibony N., Fonctions holomorphes bornées sur la boule unité de C_n, *Inventiones Math.* **67** (1982), pp. 213–222.
[16] Helson H., and Szegö G., A Problem in Prediction Theory, *Ann. Math. Pura Appl.* **51** (1960), pp. 107–138.
[17] Hörmander L., Generators for some Rings of Analytic Functions, *Bull. Amer. Math. Soc.* **73** (1967), pp. 943–949.
[18] Hunt R. A., Muckenhoupt B. and Wheeden R. L., Weighted Norm Inequalities for the Conjugate Function and Hilbert Transform, *Trans. Amer. Math. Soc.* **176** (1973), pp. 227–251.
[19] Janson S., Characterization of H^1 by Singular Integral Transforms on Martingales and $\boldsymbol{R}^n$, *Math. Scand.* **41** (1977), pp. 140–152.
[20] Jones P. W., Carleson Measures and the Fefferman–Stein Decomposition of BMO ($\boldsymbol{R}$), *Annals of Math.* **111** (1980), pp. 197–208.
[21] Jones P. W., Factorization of A_p Weights, *Annals of Math.* **111** (1980), pp. 511–530.
[22] Jones P. W., L^∞ Estimates for the $\bar{\partial}$ Problem in a Half-Plane, *Acta Math.*, to appear.
[23] Jones P. W., Interpolation Between Hardy Spaces. In: *Conference on harmonic analysis in honor of Antoni Zygmund*, Vol. 2, Wadsworth Inc., Belmont, California, 1983, pp. 437–451.
[24] Jones P. W., *BMO and the Banach Space Approximation Problem*, preprint.
[25] Jones P. W. and Marshall D. E., to appear.
[26] Løw E., A Construction of Inner Functions on the Unit Ball in C_p, *Inventiones Math.* **67** (1982), pp. 223–229.
[27] Maurey B., Isomorphismes entre espaces H_1, *Acta Math.* **145** (1980), pp. 79–120.
[28] Pommerenke Ch., On the Green's Function of Fuchsian Groups, *Ann. Acad. Sci. Fenn.* **2** (1976), pp. 409–427.
[29] Rubio de Francia J. L., Factorization and Extrapolation of Weights, *Bull. Amer. Math. Soc.* **7** (1982), pp. 393–395.
[30] Stein E. M. and Weiss G., On the Theory of Harmonic Functions of Several Variables, *Acta Math.* **103** (1960), pp. 25–62.
[31] Uchiyama A., A Constructive Proof of the Fefferman–Stein Decomposition of BMO ($\boldsymbol{R}^n$), *Acta Math.* **148** (1982), pp. 215–241.
[32] Uchiyama A., *The Fefferman–Stein Decomposition of Smooth Functions and its Applications to H^p ($\boldsymbol{R}^n$)*, thesis, University of Chicago, 1982.
[33] Uchiyama A., A Constructive Proof of the Fefferman–Stein Decomposition of BMO on Simple Martingales. In: *Conference on harmonic analysis in honor of Antoni Zygmund*, Vol. 2, Wadsworth Inc., Belmont, California, 1983, pp. 495–505.
[34] Varopoulos N. Th., BMO Functions and the $\bar{\partial}$ Equation, *Pacific J. Math.* **71** (1977), pp. 221–273.
[35] Widom H., The Maximum Principle for Multiple-Valued Analytic Functions, *Acta Math.* **126** (1971), pp. 63–81.
[36] Wojtaszczyk P., The Franklin System is an Unconditional Basis in H_1, *Arkiv för Mat.* **20** (1982), pp. 293–300.
[37] Wojtaszczyk P., Hardy Spaces on the Complex Ball are Isomorphic to Hardy Spaces on the Disc, $1 \leqslant p \leqslant \infty$, *Annals of Math.*, to appear.

INSTITUT MITTAG-LEFFLER
18262 DJURSHOLM, SWEDEN

Proceedings of the International Congress of Mathematicians
August 16–24, 1983, Warszawa

С. И. ПИНЧУК

Аналитическое продолжение отображений и задачи голоморфной эквивалентности в C^n

Настоящий обзор посвящен проблеме голоморфной эквивалентности областей в пространстве C^n. Задача состоит в следующем: как по заданным областям $D, G \subset C^n$ определить, существует биголоморфное отображение $f: D \to G$ или нет.

1. Результаты отрицательного характера

Отличие многомерной задачи голоморфной эквивалентности от соответствующей одномерной начинается с того, что в C^n при $n > 1$ нет аналогов классической теоремы Римана об отображении односвязной области на круг. Еще А. Пуанкаре показал, что такие простейшие области в C^2, как шар и бикруг, голоморфно не эквивалентны. Обобщениями этого факта являются следующие недавние результаты о том, что ограниченную область в C^n с гладкой границей нельзя биголоморфно отобразить на

а) аналитический полиэдр с кусочно-гладкой границей (Г. М. Хенкин, [14]);

б) ограниченную псевдовыпуклую область с кусочно-гладкой границей (С. И. Пинчук, [10]);

в) голоморфное расслоение, у которого база и слой имеют положительную размерность (Хаклъберри и Ормсби, [15]).

Эти результаты показывают, что голоморфным отображениям многомерных областей присуща бо́льшая жесткость (что вполне понятно, учитывая переопределенность системы уравнений Коши–Римана), и что голоморфная эквивалентность двух случайным образом выбранных областей из C^n является скорее исключением, чем правилом. Такой вывод еще более подтверждается результатом Бернса, Шнайдера и Уэллса [4] о том, что в пространстве функций, определяющих строго псевдовыпуклые области в C^n, областям, биголоморфно неэквивалентным произвольной фиксированной области соответствует всюду плотное множество второй категории.

2. Соответствие границ

Остановимся более подробно на проблеме голоморфной эквивалентности в классе строго псевдовыпуклых областей (напомним, что ограниченная область $D \subset C^n$ с границей класса C^2 называется *строго псевдовыпуклой*, если в окрестности каждой своей точки граница ∂D локально биголоморфно эквивалентна строго выпуклой гиперповерхности). Еще одна специфика многомерного случая состоит в том, что на границе такой области всякая CR-функция (т.е. функция, удовлетворяющая на ∂D касательным условиям Коши–Римана) является следом функции, голоморфной в D. Это означает, что проблема голоморфной эквивалентности строго псевдовыпуклых областей равносильна проблеме CR-эквивалентности их границ. Однако без дополнительной информации о гладкости CR-отображений, а тем более непрерывности, последняя проблема будет ничуть не проще. Поэтому важную роль играют вопросы граничного поведения отображений, вопросы соответствия границ. В этом направлении получены следующие результаты.

Пусть $f\colon D \to G$ — биголоморфное отображение, а области $D, G \subset C^n$ строго псевдовыпуклы. Тогда

а) *f непрерывно продолжается в $\bar{D}$;*

б) *если границы $\partial D, \partial G$ гладкие (класса C^∞), то f также будет гладким (C^∞) в $\bar{D}$;*

в) *если границы $\partial D, \partial G$ вещественно-аналитичны, то f аналитически продолжается в окрестность замыкания $\bar{D}$.*

Эти утверждения допускают и локальные формулировки.

Утверждение а) доказывается сравнительно элементарно при помощи весьма грубых оценок поведения метрики Каратеодори около границ строго псевдовыпуклых областей. Принадлежит оно Г. А. Маргулису [7]. Утверждение б) впервые получено Феферманом [11] в результате очень сложного и тонкого анализа граничного поведения метрики Бергмана. Оно подводит твердую базу под старые результаты Пуанкаре, Сегре, Картана и сравнительно недавние результаты Танаки [13], Мозера и Черна [16], в которых найдены различные локальные CR-инварианты вещественных гиперповерхностей с невырожденной формой Леви. С точки зрения проблемы голоморфной эквивалентности областей эти результаты также имеют отрицательный характер, так как по существу устанавливают только необходимые, но не достаточные условия CR-эквивалентности границ.

Утверждение в) получено в [8] (см. также [6], [5]). Оно сводит задачу о голоморфной эквивалентности областей к задаче о голоморфной эквивалентности их границ, которые являются строго псевдовыпуклыми вещественно-аналитическими (сокращенно АСПВ) гиперповерхностями. Голоморфные преобразования АСПВ гиперповерхностей подробно изучались в работах [13], [16]. В них найдены необходимые и достаточные условия для того, чтобы две такие гиперповерхности были локально биголоморфно эквивалентны. Однако эффективного решения проблемы о глобальной эквивалентности АСПВ гиперповерхностей они не давали. Остановимся кратко на результатах Мозера, полученных в работе [16] и имеющих отношение к рассматриваемой задаче.

3. Теория Мозера

Простейшим примером АСПВ гиперповерхности является сфера $S^{2n-1} = \{z \in C^n\colon |z| = 1\}$. С помощью биголоморфного отображения

$$('z, z_n) \to \left(\frac{'z}{1-z_n}, \frac{i(1+z_n)}{1-z_n}\right),$$

где $'z = (z_1, \ldots, z_{n-1})$, сфера преобразуется в поверхность

$$y_n = |'z|^2. \tag{1}$$

Пусть теперь $\Gamma \subset C^n$ $(n > 1)$ — произвольная АСПВ гиперповерхность и $0 \in \Gamma$. Мозером доказано, что после подходящего биголоморфного преобразования уравнение Γ в окрестности начала координат может быть приведено к следующему виду, напоминающему (1),

$$y_n = |'z|^2 + \sum_{k,l \geqslant 2} F_{kl}('z, '\bar{z}, x_n), \tag{2}$$

где F_{kl} — однородные многочлены степени k, l по $'z$, $'\bar{z}$ соответственно, коэффициенты которых аналитически зависят от x_n. При этом коэффициенты F_{22}, F_{32}, F_{33} удовлетворяют некоторым дополнительным условиям (вид которых для нас несущественен). В этом случае уравнение (2) называется *нормальной формой гиперповерхности Γ*. Две гиперповерхности Γ, $\Lambda \subset C^n$ с отмеченными точками $p \in \Gamma$, $q \in \Lambda$

локально голоморфно эквивалентны тогда и только тогда, когда они имеют в этих точках одинаковые наборы нормальных форм. АСПВ гиперповерхность можно привести к нормальной форме в заданной точке многими способами. Однако оказывается (и это один из основных результатов Мозера), что все нормальные формы АСПВ гиперповерхности в точке зависят лишь от конечного числа параметров, имеющих простой геометрический смысл. Впрочем в точках общего положения—так называемых *омбилических точках*—эта неоднозначность может быть практически ликвидирована за счет дополнительных ограничений на коэффициенты в (2) (см. [4]). В то же время уравнение (2) зависит от бесконечного числа параметров. Таким образом, множество всех классов локально эквивалентных друг другу ростков АСПВ гиперповерхностей также зависит от бесконечного числа параметров. Важным, но очень частным, случаем АСПВ гиперповерхностей являются так называемые сферические гиперповерхности — те, которые в окрестности каждой своей точки локально эквивалентны сфере.

4. Аналитическое продолжение и глобальная эквивалентность АСПВ гиперповерхностей

В 1974 г. Александер [1] получил удивительный результат, не имеющий аналогов в одномерном случае: *всякое биголоморфное отображение, преобразующее кусок сферы в кусок сферы, продолжается до голоморфного автоморфизма всего шара.* В связи с этим результатом естественно возникла следующая гипотеза. Пусть Γ, $\Lambda \subset C^n$ — компактные АСПВ гиперповерхности и f — биголоморфное отображение, которое преобразует кусок гиперповерхности Γ в кусок гиперповерхности Λ. Тогда f должно аналитически продолжаться по любому пути на Γ.

Эта гипотеза была доказана в [8] для случая, когда Λ есть сфера. Однако следующий пример, принадлежащий Бёрнсу и Шнайдеру [3] (и построенный ими для других целей), показывает, что в общем случае она неверна. Действительно, пусть

$$\Gamma = \{z \in C^2\colon\ y_2 = |z_1|^2\},$$
$$\Lambda = \{z \in C^2\colon\ \sin\ln|z_2| + |z_1|^2 = 0,\ e^{-\pi} \leqslant |z_2| \leqslant 1\}.$$

Тогда отображение $(z_1, z_2) \to (z_1/\sqrt{z_2}, e^{i\ln z_2})$ переводит $\Gamma\setminus\{0\}$ в Λ, но не продолжается в точку $z = 0$ (то, что поверхность Γ некомпактна, несущественно, так как она эквивалентна сфере). В рассмотренном примере Λ — неодносвязная компактная сферическая поверхность.

Именно это обстоятельство служит препятствием к продолжению f в точку $z = 0$. И как на первый взгляд ни странно, случай сферических поверхностей является исключительным в том смысле, что только для них приведенная гипотеза перестает быть верной. Имеет место следующий результат.

Теорема 1 [9]. *Пусть Γ, $\Lambda \subset C^n$ — несферические связные* АСПВ *гиперповерхности, причем Λ компактна. Пусть существует непостоянное голоморфное отображение f некоторого открытого подмножества $\Gamma_1 \subset \Gamma$ в Λ. Тогда f аналитически продолжается по любому пути на Γ как локально биголоморфное отображение.*

Полное доказательство теоремы 1 довольно длинно и существенно опирается на свойства нормальных форм и метрики Фефермана [12]. Не имея здесь возможности даже кратко его изложить, отметим лишь одно важное промежуточное утверждение, которое в первом приближении состоит в том, что в условиях теоремы 1 f может осуществлять большие растяжения только в направлениях (в образе), близких к комплексным касательным. При этом весьма полезным является следующее наблюдение. Пусть уравнение Γ в окрестности нуля приведено к нормальной форме (2). Рассмотрим биголоморфное преобразование

$$('z, z_n) \to \left(\frac{1}{\sqrt{\delta}}\, U('z),\ \frac{1}{\delta} z_n\right), \tag{3}$$

где $U\colon C^{n-1} \to C^{n-1}$ унитарно, а $\delta > 0$. Тогда образ Λ поверхности Γ определяется уравнением

$$y_n - |'z|^2 = \frac{1}{\delta} \sum_{k,l \geqslant 2} F_{kl}\left(\sqrt{\delta}\, U^{-1}('z),\ \sqrt{\delta}\, \overline{U^{-1}('z)},\ \delta x_n\right), \tag{4}$$

которое также имеет нормальную форму. Так как k, $l \geqslant 2$, то все члены в правой части (4) стремятся к нулю при $\delta \to 0$. Отображение (3) осуществляет при $\delta \to 0$ сколь угодно большие растяжения по всем направлениям, однако это происходит за счет того, что в пределе нарушается несферичность поверхности Λ.

Все основные результаты Мозера о нормальных формах также справедливы и для вещественно-аналитических гиперповерхностей, имеющих невырожденную форму Леви. Поэтому естественно было бы попытаться обобщить и теорему 1 на случай таких поверхностей (на комплексных многообразиях). Однако при отображениях несферических гиперповерхностей с невырожденной, но незнакоопределенной формой Леви, сколь угодно большие растяжения в некомплексных направлениях оказываются возможными. Соответствующий пример построен В. К. Белошапкой [2]. Причина этого состоит в том, что отображение (3) сохраняет нормальную форму таких поверхностей, если U — псевдоунитарное преобразование, сохраняющее соответствующую незнакоопределенную эрмитову форму. Но такое преобразование U, как и его обратное, может иметь сколь угодно большую норму. Поэтому уже нельзя делать вывод о том, что при $\delta \to 0$ правая часть (4) стремится к нулю.

Из теоремы 1 сразу получается

Теорема 2 [9]. *Для того, чтобы строго псевдовыпуклые области D, $G \subset C^n$ с односвязными вещественно-аналитическими границами были биголоморфно эквивалентны, необходимо и достаточно, чтобы нашлись точки $p \in \partial D$, $q \in \partial G$, их сколь угодно малые окрестности $U \ni p$, $V \ni q$ и биголоморфное отображение f: $U \cap \partial D \to V \cap \partial G$.*

В случае неодносвязных АСПВ границ из голоморфной эквивалентности их кусков вообще говоря уже не следует эквивалентность самих областей, так как при аналитическом продолжении вдоль границы теперь может получиться многозначное отображение. Соответствующие примеры легко строятся, но они также оказываются исключением из правила.

Назовем АСПВ гиперповерхность Γ *вполне неоднородной*, если окрестности (относительно Γ) любых двух различных точек $p_1, p_2 \in \Gamma$ голоморфно не эквивалентны, т.е. не существует биголоморфного отображения φ: $U_1 \cap \Gamma \to U_2 \cap \Gamma$ (где $U_1 \ni p_1$, $U_2 \ni p_2$) такого, что $\varphi(p_1) = p_2$. Теорема 2 остается справедливой для произвольных строго псевдовыпуклых областей D, $G \subset C^n$ с вещественно-аналитическими и вполне неоднородными границами. Действительно, такие границы очевидно несферичны. Поэтому любое биголоморфное отображение f: $U \cap \partial D \to V \cap \partial G$ в силу теоремы 1 аналитически продолжается по любому пути на ∂D, а поскольку ∂D, ∂G вполне неоднородны, это продолжение будет однозначным.

Вполне неоднородность АСПВ гиперповерхностей есть свойство общего положения в том смысле, что в пространстве функций, определяющих АСПВ гиперповерхности, вполне неоднородным гиперповерхностям соответствует всюду плотное множество второй категории. Это следует из того, что множество всех классов эквивалентных ростков АСПВ гиперповерхностей зависит от бесконечного числа параметров, тогда как размерность гиперповерхности конечна. Аналогичное утверждение для строго псевдовыпуклых гиперповерхностей класса C^∞ и их CR-отображений было получено в [4].

Литература

[1] Alexander H., Holomorphic mappings from the ball and polydisc, *Math. Ann.* **209** (1974), стр. 249–256.

[2] Белошапка В. К. Пример непродолжаемого голоморфного преобразования аналитической гиперповерхности, *Матем. заметки* **32, № 1** (1982), стр. 121–123.

[3] Burns D., Jr., and Shnider S., Spherical hypersurfaces in complex manifolds, *Invent. Math.* **33, 3** (1976), стр. 223–246.

[4] Burns D., Jr., Shnider S., and Wells R. O., Jr., Deformations of strictly pseudoconvex domains, *Invent. Math.* **46** (1978), стр. 237–253.

[5] Webster S., On the reflection principle in several complex variables, *Proc. Amer. Math. Soc.* **71** (1978), стр. 26–28.

[6] Lewy H., On the boundary behaviour of holomorphic mappings. *Accad. Naz. dei Lincei* **35**, 1977.

[7] Маргулис Г. А., *Соответствие границ при биголоморфных отображениях многомерных областей*, тезисы докладов Всес. конф. по теории функций комплексного переменного, Харьков, 1971, стр. 137–138.

[8] Пинчук С. И., Об аналитическом продолжении голоморфных отображений, *Матем. сб.* **98, № 3** (1975), стр. 416–435.

[9] Пинчук С. И., О голоморфных отображениях вещественно-аналитических гиперповерхностей, *Матем. сб.* **105, № 4** (1978), стр. 574–593.

[10] Пинчук С. И., Голоморфная неэквивалентность некоторых классов областей в C^n, *Матем. сб.* **111, № 1** (1980), стр. 67–94.

[11] Fefferman C., The Bergman kernel and biholomorphic mappings of pseudoconvex domains, *Invent. Math.* **26** (1974), стр. 1–65.

[12] Fefferman C., Monge–Ampère equations, the Bergman kernel and geometry of pseudoconvex domains, *Ann. of Math.* **103** (1976), стр. 395–416.

[13] Tanaka N., On the pseudoconformal geometry of hypersurfaces of the space of n complex variables, *Jour. Math. Soc. Japan* **14** (1962), стр. 397–429.

[14] Хенкин Г. М., Аналитический полиэдр голоморфно не эквивалентен строго псевдовыпуклой области, *Докл. АН СССР* **210, № 5** (1973), стр. 1026–1029.

[15] Huckleberry A. and Ormsby E., Non-existence of proper holomorphic maps between certain complex manifolds, *Manuscripta Math.* **26**, No. **4** (1979), стр. 371–379.

[16] Chern S. S. and Moser J. K., Real hypersurfaces in complex manifolds, *Acta Math.* **133** (1974), стр. 219–271.

БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ 40-ЛЕТИЯ ОКТЯБРЯ

Index

Aizenman, M., 1297
Ambrosetti, A., 1125
Andrianov, A. N. 465
Araki, H., 3
Arnold, V. I., 27
Arthur J. 849
Askey, R., 935

Ball, J. M., 1309
Barth, W., 783
Beilinson, A., 699
Bony, J.-M., 1133
Bourgain, J., 945
Brillinger, D. R., 1049
Brockett, R. W., 1357
Buslaev, V. S., 1149

Caffarelli, L. A., 1165
Cheng, S. Y., 533
Cherlin, G. L., 301
Chibisov, D. M., 1063
Cohen, F. R., 621
Cohen, R. L., 627

Dahlberg, B. E. J., 953
De Giorgi, E., 1175
Donaldson, S. K., 641

Erdös, P., 51
Eskin, G., 1165

Feng Kang, 1439
Figiel, T., 961
Fleming, W. H., 71
Foata, D., 1541
Fontaine, J.-M., 475
Fornæss, J. E., 791
Freedman, M. H., 647
Freudenthal, H., 1695
Fulton, W., 711

Girard, J.-Y., 307
Glowinski, R., 1455
Graham, R. L., 1555
Griess, R. L., Jr. 369
Gromov, M., 385

Harris, J., 719
Harvey, R., 797
Heath-Brown, D. R., 487
Henkin, G. M., 809
Hitchin, N. J., 541
Hooley, C., 85
Hsiang, Wu-chung, 99

Iitaka, S., 727
Iskovskih, V. A., 733
Исмагилов, Р. С., 861
Iwaniec, T., 1193

Jantzen, J. G., 393
Jones, P. W., 829
Joseph, A., 403

Karp, R., 1601
Кашин, Б. С., 977
Kasparov, G. G., 987
Katok, A., 1245
Kerckhoff, S. P., 665
Kesten, H., 1081
Khachiyan, L. G., 1569
Khovanskiĭ, A. G., 549
Klainerman, S., 1209
Knobloch, H. W., 1369
Kopell, N., 1645

Kuržanskiĭ, A. B., 1381
Kuznetsov, Yu. A., 1509

Ladyženskaya, O., 1315
Lasota, A., 1255
Lax, P. D., 119
Letichevsky, A. A., 1611
van Lint, J. H., 1579
Lin Wen-Hsiung, 679
Lions, P.-L., 1403
Loeb, P. A., 323
Lovász, L., 1591
Lusztig, G., 877

MacPherson, R. D., 213
Majda, A., 1217
Malliavin, P., 1089
Mañé, R., 1269
Mandelbrot, B. B., 1661
Mandl, P., 1097
Maslov, V. P., 139
Masser, D. W., 493
Mazur, B., 185
Meyer, Y., 1001
Micchelli, C. A., 1523
Misiurewicz, M., 1277
van Moerbeke, P., 881
Mori, S., 747
Müller, W., 565

Nirenberg, L., 15

Ogus, A., 753
Ol'šanskiĭ, A. Yu., 415
Oshima, T., 910

Parthasarathy, R., 905
Pavlov, B. S., 1011
Pełczyński, A., 237
Пинчук, С. И., 839
Pisier, G., 1027
Погорелов, А. Б., 1711
Powell, M. J. D., 1525

Ribet, K. A., 503
Ringel, C. M., 425
Rockafellar, R. T., 1419
Ruelle, D., 237

Schmidt, W. M., 515
Schoen, R. M., 575
Schwartz, J., 21
Sell, G. R., 1283
Serrin, G. B., 1717
Shaneson, J. L., 685
Shore, R. A., 337
Simon, L., 579
Siu, Yum-Tong, 287
Slisenko, A. O., 347
Soulé, C., 437
Stanley, R. P., 447
Stroock, D. W., 1107
Svirezhev, Yu. M., 1677

Takhtajan, L. A., 1331
Tarjan, R. E., 1619
Teissier, B., 763

Uhlenbeck, K. K., 585

Valiant, L. G., 1637
Venkov, A. B., 909
Vergne, M., 921
Vinberg, E. B., 593
Виро, О. Я., 603
Voiculescu, D., 1041

Waldspurger, J.-L., 525
Wall, C. T. C., 11
Watanabe, S., 1117
Woronowicz, S., 1347

Zabczyk, J., 1425
Zakharov, V. E., 1225
Zel'manov, E. I., 455
Zil'ber, B. I., 359