PROCEEDINGS OF THE
INTERNATIONAL CONGRESS OF
MATHEMATICIANS
1962

PROCEEDINGS OF THE INTERNATIONAL CONGRESS OF MATHEMATICIANS

15–22 AUGUST 1962

INSTITUT MITTAG-LEFFLER
Djursholm, Sweden

PRINTED IN SWEDEN BY

Almqvist & Wiksells

BOKTRYCKERI AKTIEBOLAG

UPPSALA 1963

PREFACE

This volume contains the official record of the International Congress of Mathematicians held in Stockholm from 15 to 22 August 1962, and the text of the addresses given by invitation of the Organizing Committee. Manuscripts were prepared for printing by Dr. V. STENSTRÖM who also read the proofs. The short communications made by members at the Congress are not included in the *Proceedings*, but the names of the communicators and the titles of their papers will be found in the section giving the scientific programme. Summaries of these contributions, if received in time, were printed in *Abstracts of Short Communications*, issued to members during the Congress.

Stockholm, June 1963.

CONTENTS

Acknowledgements ix

Scientific Programme x

Secretary's Report xxxiv

Opening Session xxxviii

Closing Session xli

The Work of the Fields Medallists[1] xliv

One-hour Addresses 1

Half-hour Addresses 219

Index 597

[1] Gårding's and Whitney's papers on the work of Hörmander and Milnor respectively are given here.

ACKNOWLEDGEMENTS

The Organizing Committee owes a great debt of gratitude to all those who have supported the Congress, either financially or by rendering services or by showing hospitality towards Congress Members. Their contribution in one way or the other has been of the utmost value for the running of the Congress.

Here are mentioned only the supporting bodies, not the many individuals whose help by word and deed is as highly appreciated.

The Swedish Government
International Mathematical Union
International Council of Scientific Unions
The City Council of Stockholm
The University of Stockholm
The Royal Institute of Technology
The Departments of Mathematics of the Universities of Lund, Stockholm and Uppsala and of the Royal Institute of Technology
The Mittag-Leffler Institute
The Post Office
The Telegraph Office

Allmänna Svenska Elektriska AB
Almqvist & Wiksells Boktryckeri AB
Bolidens Gruvaktiebolag
Facit Electronics AB
International Business Machines Svenska AB
Kockums Mekaniska Verkstads AB
Scandinavian Airlines System
Svenska Handelsbanken
AB Svenska Kullagerfabriken
Svenska Livförsäkringsbolags Förening
Telefon AB L. M. Ericsson
Trafik AB Grängesberg-Oxelösund

SCIENTIFIC PROGRAMME

One-Hour Addresses by Invitation of the Organizing Committee

Ahlfors, L. V., Teichmüller spaces.
Borel, A., Arithmetic properties of linear algebraic groups.
Church, A., Logic, arithmetic, and automata.
Dynkin, E. B. (Read by A. N. Kolmogorov), Марковские процессы и задачи анализа. (Markov processes and problems in analysis.)
Eckmann, B., Homotopy and cohomology theory.
Gel'fand, I. M. (Read by G. W. Mackey), Автоморфные функции и теория представлений. (Automorphic functions and the theory of representations.)
Grauert, H., Die Bedeutung des Levischen Problems für die analytische und algebraische Geometrie.
Henrici, P., Problems of stability and error propagation in the numerical integration of ordinary differential equations.
Kahane, J.-P., Transformées de Fourier des fonctions sommables.
Milnor, J., Topological manifolds and smooth manifolds.
Newman, M. H. A., Geometrical topology.
Nirenberg, L., Some aspects of linear and nonlinear partial differential equations.
Šafarevič, I. R., Поля алгебраических чисел. (Algebraic number fields.)
Selberg, A., Discontinuous groups and harmonic analysis.
Serre, J.-P., Géométrie algébrique.
Tits, J., Groupes simples et géométries associées.

Half-Hour Addresses by Invitation of the Organizing Committee

The Sections

1. Logic, Foundations and History.
2. Algebra and Theory of Numbers.
3. Analysis.
4. Topology and Differential Geometry.
5. Algebraic Geometry.
6. Probability and Statistics.
7. Applied Mathematics, Mathematical Physics and Numerical Analysis.
8. Education.

SECTION 1

Łoś, J., Remarks on foundations of probability. Semantical interpretations of the probability of formulas.
Scott, D., The problem of measure. (No manuscript received.)

SECTION 2

AUSLANDER, M., Modules over unramified regular local rings.
CASSELS, J. W. S., Arithmetic on an elliptic curve.
DWORK, B., A deformation theory for the zeta function of a hypersurface.
KNESER, M., Einfach zusammenhängende algebraische Gruppen in der Arithmetik.
KOSTRIKIN, A. I., Алгебры Ли и конечные группы. (Lie algebras and finite groups.)
LINNIK, JU. V., Аддитивные задачи и собственные числа модулярных операторов. (Additive problems and eigenvalues of the modular operators.)
NOVIKOV, P. S., Алгоритмические вопросы алгебры. (Algorithmic problems of algebra.) (No manuscript received.)
SUZUKI, M., A class of doubly transitive permutation groups.
TATE, J., Duality theorems in Galois cohomology over number fields.
THOMPSON, J. G., Two results about finite groups.

SECTION 3

AGMON, S., Unique continuation and lower bounds for solutions of abstract differential equations.
ANDREOTTI, A., Complex pseudoconcave spaces and automorphic functions.
BAILY, W. L. JR., On the moduli of Abelian varieties with multiplications from an order in a totally real number field.
CARLESON, L., Interpolations by bounded analytic functions and the corona problem.
CHOQUET, G., Les cônes convexes faiblement complets dans l'analyse.
COHEN, P. J., Idempotent measures and homomorphisms of group algebras.
EHRENPREIS, L., Conditionally convergent functional integrals and partial differential equations.
HÖRMANDER, L., Existence, uniqueness and regularity of solutions of linear differential equations.
JENKINS, J. A., On normalization in the general coefficient theorem.
KOLMOGOROV, A. N., Различные подходы к оценке трудности приближенного задания и вычисления функций. (Various approaches to estimating the difficulty of approximate representation and calculation of functions.)
KURANISHI, M., On deformations of compact complex structures.
LERAY, J., Prolongement de la transformation de Laplace.
MALLIAVIN, P., Ensembles de résolution spectrale.
NACHBIN, L., Résultats récents et problèmes de nature algébrique en théorie de l'approximation.
NARASIMHAN, R., The Levi problem in the theory of functions of several complex variables.
PJATECKIĬ-ŠAPIRO, I. I. (Read by I. R. ŠAFAREVIČ), Обпасти типа верхней полуплоскости в теории функций многих комплексных переменных. (Regions generalizing the upper half plane in the theory of functions of several complex variables.)
PLIŚ, A., Unique continuation theorems for solutions of partial differential equations.

ŠILOV, G. E. (Read by B. MALGRANGE)., Корректные задачи для линейных уравнений в частных производных с постоянными коэффициентами в вещественном полупространстве. (Correctly posed problems for linear constant coefficient partial differential equations in the real half-space.)

STAMPACCHIA, G., Second order elliptic equations and boundary value problems.

STEIN, E. M., Conjugate harmonic functions in several variables.

SZÖKEFALVI-NAGY, B., The "outer functions" and their role in functional calculus.

WERMER, J., Maximal ideal spaces.

SECTION 4

ADAMS, J. F., Applications of the Grothendieck-Atiyah-Hirzebruch functor $K(X)$.

ATIYAH, M. F., The Grothendieck ring in geometry and topology.

BERGER, M., Les variétés riemanniennes dont la courbure satisfait certaines conditions.

BING, R. H., Embedding surfaces in 3-manifolds.

DOLD, A., Partitions of unity in the theory of fibration.

FLOYD, E. E., Some connections between cobordism and transformation groups.

GANEA, T., On some numerical homotopy invariants.

KATĚTOV, M., Allgemeine Stetigkeitsstrukturen.

KERVAIRE, M. A., Differentiable structures on spheres and homotopy.

POÉNARU, V., Produits cartésiens de variétés différentielles par un disque.

SMALE, S., Dynamical systems and the topological conjugacy problem for diffeomorphisms.

SMIRNOV, JU., Некоторые вопросы равномерной топологии. (Some questions of uniform topology.)

STALLINGS, J. R. JR., Topological unknottedness of certain spheres. (No manuscript received.)

WHITEHEAD, G. W., Some aspects of stable homotopy theory.

SECTION 5

HIRONAKA, H., On resolution of singularities (characteristic zero).

IGUSA, J., Structure theorems of modular varieties.

MUMFORD, D., Projective invariants of projective structures and applications.

SECTION 6

HUNT, G. A., Transformation of Markoff processes.

ITÔ, K., The Brownian motion and tensor fields on Riemannian manifold.

SINAĬ, JA. G., Вероятностные идеи в эргодической теории. (Probabilistic ideas in ergodic theory.)

SECTION 7

GRAD, H., Mathematical problems in magneto-fluid dynamics and plasma physics.

MOSER, J., New results on the stability of periodic motions.

WIGHTMAN, A. S., Some results on the structure of relativistic quantum field theory.

SHORT COMMUNICATIONS

GIVEN BY SECTIONS WITH THE AUTHORS ARRANGED ALPHABETICALLY

Section 1—Logic, Foundations and History

AABOE, ASGER H., *Pre-Ptolemaic Greek astronomy.*

BEČVÁR, JIŘÍ, *Finite and combinatorial automata.*

BOONE, WILLIAM W., *Thue system with word problem of any preassigned recursively enumerable degree of unsolvability.*

DE BOUVÈRE, K. L., *On formal ambiguity.*

VAN DALEN, D., *Some problems in intuitionistic affine geometry.*

DAVIS, MARTIN D., *Quantification theory as a free variable calculus.*

DIAMOND, ROBERT J., *Each and all.*

DIJKMAN, JAC. G., *Intuitionism and probability theory.*

EISELE-HALPERN, CAROLYN, *Fermatian inference and De Morgan's syllogism of transposed quantity in Peirce's logic of science.*

GANDY, ROBIN O., *Recursive functionals of types 3 and 4.*

GINSBURG, SEYMOUR and ROSE, GENE F., *Some recursively unsolvable problems in ALGOL-like languages.*

GOODSTEIN, REUBEN L., *A recursive lattice.*

GRELL, H., and BUDACH, L., *Erweiterungstheorie von eindimensionalen Noetherschen Präschemata.*

HARZHEIM, EGBERT, *Some theorems on lexicographically ordered sets.*

HIJAB, WASFI A., *A logical characterization of rigid language games.*

KALMÁR, LÁSZLÓ, *Some heuristical ideas about a qualitative theory of information.*

LEE, CHI-YUAN, *A self-reproducing property of Turing machines.*

LÉVY, AZRIEL, *The interdependence of some consequences of the axiom of choice.*

MAL'CEV, A. I., О теориях I-й ступени некоторых классов групп и колец. (*Theories of the first order of some classes of groups and rings.*)

MOISIL, G. C., *La logique à trois valeurs et ses applications.*

MÜLLER, G. H., *Charakterisierung einer Klasse von rekursiven Funktionen.*

NERODE, ANIL, *Inequalities in recursive equivalence types.*

OBERSCHELP, ARNOLD, *Klassen als "Urelemente" in der Mengenlehre.*

PORTE, JEAN, *An associativity lemma.*

ROBINSON, T. T., *The VH-property and constructive aspects of Harrop's theorems.*

RÖDDING, DIETER, *Darstellungen der (im Kalmárschen Sinne) elementaren Funktionen.*

SACKS, GERALD E., *Recursively enumerable degrees.*

SALOMAA, ARTO, *A completeness criterion for sets of many-valued truth-functions.*

SCHRÖTER, K. W., *Beweis der Church'schen Hypothese betreffend den Begriff der Entscheidbarkeit.*

SHOENFIELD, JOSEPH R., *A separation theorem in the predicate calculus.*

SKOLEM, THORALF A., *A theorem on recursively enumerable sets.*

SMULLYAN, RAYMOND M., *Analytic natural deduction.*

STABLER, E. RUSSELL, *Mathematics and the "central problem of our time".*

STOJAKOVIĆ, MIRKO, *Über die Entscheidbarkeit des Vierfarbenproblems.*

TAĬMANOV, A. D., Характеристика элементарных классов моделей. (*The characteristic of elementary classes of models.*)

TAMARI, DOV, *The associativity problem for finite monoids is unsolvable and equivalent to the word problem for finitely presented groups (WPG).*

TARSKI, ALFRED, *Undecidability of the elementary theory of commutative semigroups.*

TURQUETTE, ATWELL R., *Independent axioms for infinite-valued logic.*

VIOLA, TULLIO, *Un théorème sur les treillis.*

Section 2—Algebra and Theory of Numbers.

ABBOTT, JAMES C., *The imbedding of an implication algebra in a Boolean algebra.*

ADJAN, S. I., Правило сокращения в конечноопределенных полугруппах. (*The cancellation law in finitely defined semigroups.*)

AIGNER, ALEXANDER, *Folgen der Art* ar^n+b, *welche nur teilbare Zahlen liefern.*

ALLING, NORMAN L., *Valuation theory applied to function rings.*

ALMEIDA COSTA, A., *Sur les* μ*-demi-anneaux.*

ALMERING, J. H. J., *On a system of Diophantine equations, connected with a geometric problem.*

AMIR-MOÉZ, ALI R., *Singular values of sets of linear transformations.*

AMITSUR, SHIMSHON A., *Rational identities and foundation of geometry.*

ANDRUSHKIW, JOSEPH W., *A characterization of rings orderable in a non-Archimedean way.*

APOSTOL, TOM M., *Bohr's equivalence relation for Dirichlet series.*

BASS, HYMAN, *Projective modules and vector bundles.*

BAUR, ARNOLD, *Rationale Punkte auf Kurven dritter Ordnung.*

BECKEN, SIGRID, *Eine gruppentheoretische Kennzeichnung von orthogonalen Gruppen.*

BEHRENS, ERNST-AUGUST, *Die Halbgruppe der Ideale einer Algebra mit distributivem Idealverband.*

BERNSTEIN, LEON, *Zur Lösung der Diophantischen Gleichung* $1=1/x_1+1/x_2+\ldots+1/x_n$.

BLOCK, RICHARD E., and ZASSENHAUS, HANS, *The Lie algebras with a non-degenerate trace form.*

BOHUN-CHUDYNIV, VOLODYMYR, *On loops represented by two-sided triplets.*

BOHUN-CHUDYNIV, VOLODYMYR, *On a transformator of multiplicative systems.*

BOSTOCK, FRED A., *A generalisation of the Divinsky radical.*

BRAUER, RICHARD, *On finite groups of even order.*

BRIGGS, WILLIAM E., *Prime-like sequences generated by a sieve process.*

BRUCK, RICHARD H., *On the completion of finite partial planes.*

BRUN, VIGGO, *Euclidean algorithms and musical theory.*

CARTER, ROGER W., *Nilpotent self-normalizing subgroups of certain insoluble groups.*

CHALK, J. H. H., *Small solutions of some congruences.*

CHAWLA, LAL MUHAMMAD, *The perfect square congruence over integers.*

CHEEMA, MOHINDAR SINGH, *Vector partitions.*

CHOWLA, PAROMITA, *Relations between classnumbers of real and imaginary quadratic fields.*

CHOWLA, SARVADAMAN, *Elementary remarks on the zeta function of an algebraic variety.*

COHN, HARVEY, *On modular invariants for some fundamental domains in two complex variables.*

COHN, RICHARD M., *Linked ideals.*

COXETER, H. S. M., *The symmetry group of the regular complex polygon* $p_1\{q\}p_2$.

CRAWLEY, PETER, and JÓNSSON, BJARNI, *Refinements for infinite direct decompositions of groups.*

DAVENPORT, HAROLD, *Homogeneous Diophantine equations.*

DAVID, MARCEL, *Résolution de* $x^2+y^2=z^2+m$ *par réduction matricielle.*

DAYKIN, DAVID E., *Representations of integers.*

DELANGE, HUBERT, *Sur une classe de fonctions arithmétiques.*

DIENER, KARL-HEINZ, *A theorem on free algebras.*

DLAB, VLASTIMIL, *The axiomatization of dependence relations.*

EAGON, JOHN A., and NORTHCOTT, D. G., *A generalization of the Koszul complex.*

EDGE, WILLIAM L., *An orthogonal group of order* $2^{13}\cdot 3^5\cdot 5^2\cdot 7$.

ETHERINGTON, IVOR M. H., *Totally symmetric entropic quasigroups.*

FAITH, CARL, and UTUMI, YUZO, *Quotient rings.*

FIEDLER, MIROSLAV, *On some properties of positive definite matrices.*

FRAENKEL, AVIEZRI S., *A class of transcendental numbers.*

FRANK, EVELYN, *Continued fraction expansions for binomial quadratic surds.*

FRANKEL, EDWARD T., *A chessboard analog to Moessner's theorem on powers of integers.*

FREUDENTHAL, H., *Real types and real representations of semi-simple Lie algebras.*

FRIED, ERVIN, *Isomorphism theorems for abstract algebras.*

FRONTERA, BARTOLOMÉ, *A necessary condition for the reducibility of finite lattices.*

FULTON, CURTIS M., *Vector space axioms for geometry.*

GAGNON, FRANÇOIS-PIERRE, *Demi-corps et corpoïdes.*

GILMORE, P. C., and HOFFMAN, A. J., *Characterizations of comparability and interval graphs.*

GŁADYSZ, S., *Konvexe Mengen in topologischen Gruppen.*

GLODEN, ALBERT, *Une méthode pour construire des sytèmes multigrades normaux.*

GODWIN, HERBERT J., *Relations between cubic and quartic fields.*

GONSHOR, HARRY, *Polyploidy algebras with multiple allelles.*

GORDON, BASIL, *Some Ramanujan-like continued fractions.*

GROSSWALD, EMIL, *On a theorem of Petersson and Meinardus.*

GUÉRINDON, JEAN, *Sur la topologie spectrale des anneaux commutatifs.*

GUNDLACH, KARL-BERNHARD, *Die Bestimmung der Funktionen zur Hilbertschen Modulgruppe des Zahlkörpers* $P(\sqrt{5})$.
GUT, MAX, *Abschätzungen für die Klassenzahlen der quadratischen Körper.*
HALL, MARSHALL, JR. *Note on the Mathieu group* M_{12}.
HALPERIN, ISRAEL, *Elementary divisors in von Neumann rings.*
HANANI, HAIM, *Tactical configurations.*
HEINEKEN, HERMANN, *Probleme in Gruppen mit Engelbedingung.*
HELD, DIETER, *Engelsche Elemente in endlichen Gruppen.*
HERRMANN, OSKAR, *Über die Verteilung von Gitterpunkten im hyperbolischen Raum.*
HONDA, KIN-YA, *On Abelian p-groups with special principal systems.*
HOWIE, JOHN M., *Free products of semigroups with amalgamation.*
HUSSAIN, S. M., *Evaluation of a more general form of the determinant of 11th order.*
ISBELL, J. R., *Categories of algebras.*
ITO, NOBORU, *Some results on permutation groups of prime degree.*
IWASAWA, KENKICHI, *A class number formula for cyclotomic fields.*
IZBICKI, HERBERT, *On edge-colourings of pseudo-regular graphs.*
JOHNSEN, EUGENE C., *Matrix rational completions satisfying the* v, k, λ *incidence equation.*
JURKAT, WOLFGANG B., *An elementary proof of the prime number theorem with remainder term.*
KALMAN, JOHN A., *A class of non-distributive rings.*
KANOLD, HANS-JOACHIM, *Über periodische zahlentheoretische Funktionen.*
KANTZ, GEORG, *Zerlegung unendlicher Primstellen in relativ Galoisschen Körpern.*
KAPPE, WOLFGANG, *Gruppentheoretische Eigenschaften und charakteristische Untergruppen.*
KEGEL, OTTO H., *Locally finite groups with a partition.*
KEMPERMAN, J. H. B., *Generalized distributions modulo* 1.
KIRILLOV, A. A., Унитарные представления нильпотентных групп Ли. (*Unitary representations of nil-potent Lie groups.*)
KLEE, VICTOR, *The generation of affine hulls.*
KLEISLI, H., and MARANDA, J.-M., *Satellites relative to injective structures.*
KLINGEN, HELMUT, *Fourierentwicklung von Modulformen.*
KOVÁCS, L., NEUBÜSER, J., and NEUMANN, B. H., *Endliche* $E_{p,n}$*-Gruppen.*
KRASNER, MARC, *Nombre des extensions de degré donné d'un corps p-adique.*
KRISHNAN, V. SANKRITHI, *Ordered semigroups.*
KUPISCH, HERBERT, *Symmetrische Algebren mit endlich vielen unzerlegbaren Darstellungen.*
KUYK, W., *Fixed fields under automorphism groups of purely transcendental field extensions.*
LAFON, JEAN PIERRE, *Remarque sur le lemme bilinéaire de Samuel.*
LAMBEK, JOACHIM, *Recent results on rings of quotients.*
VAN LEEUWEN, L. C. A., *Contribution to non-commutative ideal theory.*
LEHRER-ILAMED, YEHIEL, *On functions of matrices.*
LEHTI, R., *On the representation of projectivities over a skew field with dyadics.*
LEICHT, J. B., *Remarks on the axiomatic theory of additive relations.*
LINGENBERG, ROLF, *Kennzeichnung der ternären orthogonalen Gruppen.*
LIPSCHUTZ, SEYMOUR, *On matrices associated with braids.*
LOONSTRA, F., *Subgroups of direct products.*

ŁOŚ, JERZY, *On homomorphisms of commutative semigroups.*
McQUILLAN, DONALD L., *Some structure theorems for SL (2, n).*
MARCZEWSKI, EDWARD, *Remarques sur les algèbres homogènes.*
MENDELSOHN, NATHAN S., *Non negative matrices.*
MILLER, D. D., *Matrices over commutative rings.*
MOORE, CHARLES N., *On the frequency of prime patterns.*
MORAN, SIEGFRIED, *Unrestricted products.*
MORDELL, LOUIS JOEL, *On Lerch's class number formulae for binary quadratic forms.*
MUWAFI, AMIN A., *A quadratic Diophantine equation.*
NEWMAN, MICHAEL FREDERICK, *Embedding theorems for finite p-groups.*
NEWMAN, MORRIS, *Diagonal quadratic forms.*
NIVEN, IVAN, *On the covering of lattice points in 3-space.*
OPPENHEIM, SIR ALEXANDER, *The rational integral solution of the equation* $x^3+y^3=u^3+v^3$ *and allied Diophantine equations.*
OSBORN, J. MARSHALL, *Power-associative division algebras.*
PAPANGHELOU, FREDOS, *On the order topologies of an Abelian lattice-group.*
PAVLOVIĆ, S., *Sur les nombres d'Euler généralisés.*
PENNING, CHRISTIAAN J., *Rings with prescribed idempotents and tensor products.*
PEREMANS, W., *Domaine and range homomorphisms.*
PICCARD, SOPHIE, *Les groupes fondamentaux et les groupes quasi libres.*
PONTING, FREDERICK W., *Partitions and condensation formulae for S-functions.*
POPKEN, JAN, *Algebraic dependence of arithmetic functions.*
POSNER, EDWARD C., and RUMSEY, HOWARD C., *Polynomials which divide infinitely many trinomials.*
PRESTON, GORDON BAMFORD, *The lattice of congruences on a completely 0-simple semigroup.*
PUMPLÜN, DIETER, *Über die Teilbarkeit der Klassenzahl gewisser imaginärquadratischer Zahlkörper durch Potenzen von 2.*
PUPPE, DIETER, *Korrespondenzen in abelschen Kategorien.*
QUERRÉ, J., *r-systèmes d'idéaux d'un demi-groupe.*
RADO, RICHARD, *A theorem on vector spaces.*
REINER, IRVING, and HELLER, ALEX, *Indecomposable integral representations of finite groups.*
ROBINSON, RAPHAEL M., *Intervals containing infinitely many sets of conjugate algebraic units.*
ROOS, JAN-ERIK I., *Derived functors of infinite products and projective objects in Abelian categories.*
ROTKIEWICZ, ANDRZEJ, *On Lucas numbers with two intrinsic prime divisors.*
RYSER, H. J., and FULKERSON, D. R., *The* α*-width of a* (0,1)*-matrix.*
SANDS, ARTHUR D., *On the factorization of Abelian groups.*
SCHIEK, HELMUT, *Das Adjunktionsproblem der Gruppentheorie.*
SCHINZEL, ANDRZEJ M., *On the arithmetic of polynomials and some related problems.*
SCHOENEBERG, BRUNO, *Eisensteinsche Reihen von Primzahlstufe.*
SCHWARZ, ŠTEFAN, *Measures on compact semigroups.*
SCHWERDTFEGER, HANS, *A special class of Frobenius groups.*
SEDLÁČEK, J., *Rationale Punkte im* E_2 *und* E_3.
SELENIUS, CLAS-OLOF, *Eine endgültige kettenbruchtheoretische Erklärung der*

alten indischen zyklischen Methode zur Lösung der Bhaskara-Gleichung $x^2 - Dy^2 = 1$ (*Pellsche Gleichung*).

SERES, IVAN (Read by P. ERDÖS), *Irreducibility of any polynomial.*

SHANKS, DANIEL, *An inductive formulation of the Riemann hypothesis.*

SIERPIŃSKI, WACŁAW, *Sur les nombres de la forme* $a^{2^n} + 1$.

SKLAR, ABE, *The algebra of operators.*

SOLOMON, LOUIS, *The invariants of finite reflection groups.*

SPRINGER, T. A., *Twisted composition algebras.*

STRAUS, ERNST G., *A remark on the p-adic Roth theorem.*

SZENDREI, JÁNOS, *Über die Quasi-Ideale von Ringen.*

TAMASCHKE, OLAF, *Zur Theorie der Permutationsgruppen mit regulärer Untergruppe.*

TEH, HOON HENG, *Delta sum of partially ordered groups.*

THEDY, A., *Computation of the groups of automorphisms of the cone of positive hermitian forms.*

THIERRIN, GABRIEL, *Une caractérisation des groupes d'ordre premier.*

TURÁN, PAUL, *A comparative theory of prime numbers.*

WARD, MORGAN, *p-adic linear recurrences.*

WATSON, G. L., *The class-number of a positive quadratic form.*

WEICHSEL, PAUL M., *On critical p-groups.*

WELLS, MARK BRIMHALL, *Bit parity of multiples of primes.*

WIELANDT, HELMUT W., *Bedingungen für die Konjugiertheit von Untergruppen endlicher Gruppen.*

WUYTACK, FRANS A., *General spin operators.*

ZELINSKY, DANIEL and ROSENBERG, ALEX, *Amitsur's complex for inseparable fields.*

Section 3—Analysis

ACZÉL, JÁNOS, *Some recent results and problems on functional equations.*

AFRIAT, SYDNEY N., *On the irreflexivity of a relation associated with a positive vector function* $x = f(u)$ $(u'x = 1)$.

AHLBERG, J. H., and NILSON, E. N., *Convergence properties of the spline fit.*

AKUTOWICZ, EDWIN J., *On spectral approximation of sequences.*

ALBRYCHT, JERZY, L^p*-spaces, with mixed norm, for finitely additive set functions.*

ALEXIEWICZ, A., *Two-norm spaces.*

ALPÁR, LÁSZLÓ, *Sur certaines transformées des séries de puissance absolument convergentes sur la frontière de leur cercle de convergence.*

ALEKSANDRJAN, R. A., О некоторых системах типа С. Л. Соболева и об однородных краевых задачах для дифференциальных операторов с индефинитной квадратичной формой. (*On some systems of S. L. Sobolev's type and on homogeneous boundary problems for differential operators with indefinite quadratic form.*)

ANTOSIEWICZ, H. A., *Maximal growth solutions of linear differential equations.*

ARNOL'D, V. I., Теория возмущений и проблема устойчивости планетных систем. (*Perturbation theory and the problem of stability for planetary systems.*)

ARSOVE, MAYNARD G., *Radial limit sets for subharmonic functions.*

AUBERT, KARL E., *Convex ideals in ordered group algebras.*

AVANISSIAN, VAZGAIN, *Fonctions entières de plusieurs variables á croissance très lente.*

AZIZ, A. K., and DIAZ, J. B., *On a mean value theorem of the differential calculus of vector valued functions, and uniqueness theorems for ordinary differential equations in a linear normed space.*

BARNETT, I. A., and MENDEL, C. W., *On a system of non-linear differential equations.*

BARRY, PATRICK D., *The minimum modulus of integral functions.*

BEHNERT-SMIRNOV, K., *Eine notwendige und hinreichende Bedingung der gleichgradigen Stetigkeit von zweimal differenzierbaren Funktionenmengen.*

BERGMAN, STEFAN, *On meromorphic functions of several complex variables.*

BICADZE, A. V., Об уравнениях смешанного типа с тремя независимыми переменными. (*On equations of mixed type in three independent variables.*)

BOGNÁR, JÁNOS, *On the existence of square-roots of J-selfadjoint operators.*

BOJANIC, R., *Application of Bergman operators to the theory of equations of mixed type.*

BOSANQUET, LANCELOT S., *A convexity theorem for fractional derivatives.*

BROUSSE, PIERRE, *Problèmes elliptiques mixtes du quatrième ordre.*

BUCK, R. CREIGHTON, *Global solutions of differential equations.*

BUREAU, FLORENT J., *Sur les équations différentielles dont l'intégrale générale est a points critiques fixes.*

BUTZER, P. L., *Approximation problems for solutions of partial differential equations.*

CALABI, EUGENIO, *Operational calculus in several variables.*

CIMMINO, GIANFRANCO, *Sur les traces des solutions des équations différentielles linéaires à coefficients constants.*

CIVIN, PAUL, *Maximal ideals in the second conjugate algebra of a group algebra.*

COLEMAN, COURTNEY STAFFORD, *Generalized Lyapunov numbers.*

COOPER, JACOB L. B., *Fourier transforms of L^p functions and sequences.*

CROWSON, HENRY L., *Closed form solutions of a second order linear ordinary differential equation with n regular singular points.*

CUDIA, DENNIS F., *Smoothness.*

DAJOVIĆ, V., *Sur l'existence des valeurs limites d'une fonction minimale de classe $H_\delta(\delta>0)$ et d'une fonction de même classe $(\delta>1)$.*

DAMKÖHLER, WILHELM, *Über die Lösbarkeit isoperimetrischer Variationsprobleme.*

DASCALOPOULOS, D., *Über die Lösung von Systemen linearer Volterrascher Integrodifferentialgleichungen von Faltungstyp n-Ordnung.*

DAVIES, ROY OSBORNE, *Covering the plane with denumerably many curves.*

DENJOY, ARNAUD, *Les équations différentielles périodiques.*

DOLINSKY, R., *Nullstellen der Abschnitte nirgends konvergenter Potenzreihen.*

DOMAR, E. YNGVE, *Closed primary ideals in certain Banach algebras.*

DOSS, RAOUF, *On the representation of continuous functions of 2 variables by means of addition and continuous functions of 1 variable.*

DRIVER, RODNEY D., *Stability of solutions of a functional-differential equation of neutral type.*

DUREN, PETER L., *Invariant subspaces of tridiagonal operators.*

DŽRBAŠJAN, M. M., Исследование некоторых неполных систем в комплексной области. (*Investigation of some incomplete systems in a complex region.*)

EBERHARD, WALTER, *Über die Greensche Funktion der linearen Eigenwertaufgabe 4. Ordnung.*

EFIMOV, N. V., and STEČKIN, S. B. (*presented by L. D. Kudrjavčev*), Чебышевские множества в банаховых пространствах. (*Čebyšev sets in Banach spaces.*)

EICHHORN, WOLFGANG, *Zur funktionenteoretischen Behandlung partieller Differentialgleichungen.*

ELLIOTT, JOANNE, *An application of the theory of Dirichlet spaces.*

EVANS, ARWEL, *Non-absolute measures.*

EVERITT, WILLIAM NORRIE, *The circle method of Weyl applied to differential equations of arbitrary order.*

EZEILO, JAMES O. C., *A property of the phase space trajectories of a third order differential equation.*

FARKAS, MIKLÓS, *On differential geometric investigation of ordinary differential equations.*

FAURE, ROBERT, *Sur certaines solutions périodiques d'équations différentielles non linéaires.*

FELL, JAMES M. G., *The topology of induced representations of groups.*

FENYÖ, S., *Über eine Anwendung der algebraischen Derivierten.*

FIFE, PAUL C., *Schauder-type estimates for a degenerating parabolic equation.*

FINN, R., *Stationary solution of the Navier-Stoker equations.*

FOX, CHARLES, *Generalized integral transforms.*

FREUD, GÉZA, *On local differentiability of higher order of functions.*

FUCHS, WOLFGANG HEINRICH J., *On a conjecture of G. Polya concerning gap series.*

FUGLEDE, BENT, *Quasi-continuity of potentials of finite energy.*

FULLERTON, ROBERT E., *Order bases in a linear topological space.*

GARCIA DE CEVALLOS, J. F., *Construction of non-Lebesgue-measurable sets in the euclidean m-dimensional space.*

GEHRING, FREDERICK W., *Extremal length definitions for the conformal capacity of space rings.*

GEYMONAT, GIUSEPPE, *Sur le problème de Dirichlet.*

GLAESER, GEORGES, *Algèbres de fonctions composées différentiables.*

GLODEN, RAOUL-FRANÇOIS, *Propriétés des polynômes associés aux fonctions de Laguerre de 2e espèce généralisées et questions connexes.*

GOLDBERG, SEYMOUR, and THORP, EDWARD O., *On some open questions concerning strictly singular operators.*

GOLDSTEIN, A. A., and CHENEY, E. W., *Tchebycheff approximation in locally convex spaces.*

GORSKI, GEORGE, *Remark on a certain theorem of H. J. Bremermann.*

GOTTSCHLING, ERHARD, *Über einen Fundamentalbereich diskontinuierlicher Gruppen.*

GOULD, GERALD G., *A Stone-Čech-Alexandroff-type compactification.*

GREGUŠ, MICHAL, *Über die asymptotischen Eigenschaften der Lösungen der linearen Differentialgleichung dritter Ordnung.*

GÜNZLER, HANS, *Generalized Fourier series.*

GYIRES, BÉLA, *A generalization of a theorem of Szegö.*

HABETHA, KLAUS, *Über das Maximumprinzip bei Lösungen elliptisch-parabolischer Differentialgleichungen 2. Ordnung.*

HALKIN, HUBERT, *On the calculus of variations with unilateral constraints.*

HARTMAN, STANISŁAW, *Some problems in the algebra of Borel measures.*

HELGASON, SIGURDUR, *Some results in invariant theory.*
HENSTOCK, RALPH, *New integrals.*
HILDEBRANDT, STEFAN, *Über die Morrey-Calkinschen Räume.*
HILLE, EINAR, *Linear fourth order differential equations.*
HIRSCHFELD, RUDOLF A., *Self-conjugate operators in Hermite space.*
HSIANG, FU CHENG, *On the strong Riesz summability of an infinite series.*
HUBER, ALFRED, *Ein räumliches Analogon des Wiman-Heinsschen Satzes.*
HUCKEMANN, FRIEDRICH, *Extremalprobleme für schlichte konforme Abbildung eines Kreisringes.*
HUSTAD, OTTE, *Convex cones with properties related to weak local compactness.*
IBRAGIMOV, I. I., Некоторые экстремальные задачи для линейных операторов в классе целых функций конечной степени. (*Some extremal problems for linear operators in the class of entire functions of finite order.*)
ISAACS, G. L., *On fractional differences.*
JABOTINSKY, ERI, *Two-variable integral transforms.*
JEANQUARTIER, PIERRE, *Solutions élémentaires d'opérateurs différentiels paraboliques hyperboliques.*
JOHNSON, GUY, *Representation of harmonic functions in the unit circle.*
KALISCH, G. K., *Titchmarsh's convolution theorem and invariant subspaces.*
KAPPOS, D. A., *Representation of abstract L-spaces.*
KAZARINOFF, NICHOLAS D., and GOODRICH, RAYMOND F., *A derivation of the Watson transformation.*
KJELLBERG, B., *On the growth of certain entire functions.*
KOHN, JOSEPH J., *Harmonic integrals on non-compact complex manifolds.*
KOLDEN, KJELL S., *The summation formula of Euler MacLaurin and its generalization.*
KÖLZOW, D., *Der Satz von Dunform-Pettis und Zerlegungssätze für abstrakte Masse.*
KORÁNYI, ADAM, and PUKÁNSZKY, LAJOS, *Holomorphic functions with positive real part on polycylinders.*
KOREVAAR, JACOB, *Distribution proof of Wiener's Tauberian theorem.*
KRABBE, GREGERS L., *A spectral theorem for operators into an interpolation space.*
KRZYŻ, JAN, *The radius of close-to-convexity for univalent functions.*
KUCZMA, MAREK, *On the Schroeder equation.*
KUDRJAVCEV, L. D., Весовые пространства. (*Weighted spaces.*)
KUHLMANN, NORBERT, *Über die Auflösung der Singularitäten 3-dimensionaler komplexer Räume.*
KUREPA, S., *Similarity of operators.*
KUTTNER, BRIAN, *A Tauberian theorem for discontinuous Riesz means.*
LADYŽENSKAJA, O. A., Квазилинейные уравнения эллиптического и параболического типов. (*Quasi-linear equations of elliptic and parabolic types.*)
LAVRENT'EV, MICHEL A., *Sur les représentations quasi-conformes.*
LEGATOS, GERASSIMOS G., *Sur une classe de solutions d'une équation différentielle ordinaire non linéaire du second ordre.*
LEIS, R., *Eine Übertragung des Schwarzschen alternierenden Verfahrens auf Randwertprobleme der Helmholtzschen Schwingungsgleichung.*
ŁOJASIEWICZ, STANISŁAW, *Triangulation of real analytic sets.*
LORCH, LEE, and SZEGO, PETER, *Higher monotonicity properties of certain Sturm-Liouville functions.*

LORENTZ, GEORGE G., *Degree of approximation by polynomials with positive coefficients.*

LUXEMBURG., WILHELMUS A. J., *Ultrapowers in analysis.*

MACCOLL, LEROY ARCHIBALD, *Quotients of sums of exponentials of polynomials.*

MCLEOD, JOHN B., *An infinite set of non-linear differential equations associated with a transport equation.*

MCMAHON, JAMES, *Invariant volume element in C^n.*

MAGENES, ENRICO, *Sur les problèmes aux limites linéaires elliptiques.*

MAGNUS, ARNE, *A new type of continued fraction.*

MAMBRIANI, ANTONIO, *Sur l'intégration des équations aux dérivées partielles, linéaires, du 2e ordre.*

MARČENKO, V. A., Обобщенная спектральная функция. (*The generalized spectral function.*)

MARDEN, ALBERT, *A sufficient condition for the bilinear relation on open Riemann surfaces.*

MARDEN, MORRIS, *On the critical points of linear combinations of Green's function.*

MARKOVIĆ, DRAGOLJUB, *Sur la limite inférieure des modules des zéros des polynômes de deux variables.*

MARTINEAU, ANDRÉ, *Croissance d'une fonction entière de type exponentiel et supports des fonctionnelles analytiques.*

MASANI, PESI RUSTOM, *Isometric flows on Hilbert space.* (Presented in German.)

MAUDE, RONALD, *Conditions for positive Φ-capacity.*

METIVIER, MICHEL, *Sur les systèmes projectifs de mesures vectorielles.*

MILICER GRUŻEWSKA, HALINA, *Les problèmes liés avec les systèmes paraboliques d'équations.*

MITCHELL, JOSEPHINE M., *Integral theorems for harmonic vectors of three real variables.*

MITROPOL'SKIĬ, JU. A., метод интегральных иногообразий в теории нелинейных дифференциальных уравнений. (*The method of integral manifolds in the theory of non-linear differential equations.*)

MITROVIĆ, DRAGIŠA, *Le problème de Hilbert sur les valeurs limites dans un anneau normé.*

NEUNZERT, HELMUT, *Über die Forsetzung von Kontraktionen eines m-dimensionalen komplexen Vektorraumes.*

NICOLESCU, M., *Criteria for subcaloricity.*

NIKOL'SKIĬ, S. M., О граничных свойствах дифференцируемых фрнкций многих переменных. (*On boundary properties of differentiable functions of several variables.*)

NOBLE, MARK E., *Some gap theorems for Dirichlet series.*

NORGUET, FRANÇOIS, *Applications de la théorie des résidus.*

OBRECHKOFF, NIKOLA DIMITROV, *Sur les séries doubles absolument sommables par les moyennes arithmétiques.*

ODHNOFF, JAN EDVARD, *Some free boundary problems for the Laplace equation.*

OLEĬNIK, O. A., Об уравенениях пограничного слоя. (*On the equations of a boundary layer.*)

OLVER, FRANK W. J., *Bounds for remainder terms in asymptotic expansions.*

ORLICZ, WŁADYSŁAW, *On a class of modular spaces.*

OVSJANNIKOV, L. V., Применение теории групп Ли к некоторым уравнениям с частными производными. (*Application of the theory of Lie groups to some partial differential equations.*)

PAYNE, LAWRENCE E., *Inequalities for uniformly elliptic operators with application to the Cauchy problem.*

PEŁCZYŃSKI, ALEKSANDER, and ROLEWICZ, STEFAN, *Best norms with respcet to isometry groups in normed linear spaces.*

PEYERIMHOFF, ALEXANDER, *Ein Mittelwertsatz für absolute Rieszsche Summierbarkeit.*

PEYOVITCH, TADIA, *Propriétés asymptotiques d'une classe d'équations différentielles.*

PHELPS, ROBERT R., *Extremal operators are multiplicative.*

POPOV, S. BLAGOJ, *Sur les équations différentielles à coefficients polynomiaux.*

POPOVICIU, T., *Sur le reste dans certaines formules de quadrature.*

PROUSE, JOHN, *Analysis of some classical problems of propagation.*

PTÁK, V., *Critical exponents of finite dimensional Banach spaces.*

PUCCI, CARLO, *A variational problem for coefficients of elliptic equations.*

RAIMI, RALPH A., *Invariant means and matrix summability methods.*

READE, MAXWELL O., *On sections of certain Laurent series.*

REICH, EDGAR, *On a characterization of quasiconformal mappings.*

RICCI, MARIA L., and RIZZONELLI PIERANITA, *On l^1 almost-periodical function.*

ROBERTSON, MALCOLM S., *Extremal problems for analytic functions with positive real part.*

ROOS, BERNARD W., and SANGREN, WARD C., *Expansions associated with a pair of singular first-order differential equations.*

ROSENBLATT, MURRAY, *Asymptotic behaviour of eigenvalues for a class of integral equations.*

RUDIN, WALTER, *The extension problem for positive-definite functions.*

RUNCK, PAUL OTTO, *Notwendige und hinreichende Bedingungen für das Konvergenzverhalten von Interpolationspolynomen.*

RUSSELL, DENNIS CHARLES, *An inclusion relation between Riesz and Riemann summability.*

SARAFYAN, DIRAN, *General existence theorems for ordinary differential equations and their systems.*

SARIO, LEO, *Analytic mappings into Riemann surfaces.*

SCHÄFFER, JUAN J., *Linear differential equations and function spaces: variation of the equation.*

SCHWEIZER, BERTHOLD, *An inequality for the confluent hypergeometric function.*

SEIBERT, PETER, and AUSLANDER, JOSEPH, *Liapunov functions and prolongations in dynamical systems.*

SENDOV, B., and PENKOV, B., *On a metric and its applications to polynomial approximation.*

SHIELDS, ALLEN L., and RUBEL, LEE, *On bounded pointwise convergence of polynomials.*

SKOF, F., *Une famille de séries qui admettent leur cercle de convergence comme coupure et qui ne sont pas lacunaires.*

ŚLADKOWSKA, JANINA, *Bounds for analytic functions of two complex variables.*

SOBOLEV, S. L., Некоторые вопросы теории численного интегрирования и интерполяции функций многих независимых переменных. (*Quelques questions de la théorie des intégrations numériques et de l'interpolation pour les fonctions de plusieurs variables indépendantes.*)

SOBOLEV, S. L., Формулы численного интегрирования на сфере. (*Les formules de l'intégration numérique sur la surface de sphère.*)

STANKOVIĆ, BOGOLJUB, *L'équation différentielle des opérateurs de J. Mikusiński.*

STRANG, W. GILBERT, *Polynomial approximation of Bernstein type.*

SUVOROV, G. D., Топологические отображения плоских областей с переменными границами. (*Topological mappings of plane regions with variable boundaries.*)

SWANSON, CHARLES A., *Elliptic operators and small handles.*

SZÁSZ, PAUL. *On generalized quasi-step parabolas.*

SZILÁRD, KARL, *Die Analoga der ganzen rationalen Funktionen in verallgemeinerten Funktionenklassen.*

TAAM, CHOY-TAK, *Integration on locally compact Hausdorff spaces.*

TAKAHASHI, REIJI, *La série discrète dans le groupe de De Sitter.*

TANDORI, K., *Über die unbedingte Konvergenz der Orthogonalreihen.*

TEODORESCU, N. V., *Fonctions à dérivée aréolaire généralisée d'ordre $n+1$ sommable et leur représentation intégrale.*

THIELMAN, H. P., *Functional equation arising in elasticity.*

THRON, W. J., *Sequences of linear fractional transformations and continued fractions.*

TURRITTIN, HUGH L., *Reduction of ordinary differential equations to the Birkhoff canonical form.*

ULLMAN, JOSEPH L., *Tchebycheff quadrature is possible on an infinite interval.*

ULUÇAY, CENGIZ, *On Bieberbach's conjecture, Part I.*

VEKUA, I. N., *Compactness of families of solutions of elliptic differential equations.*

VINCZE, E., *Eine neuere elementare Lösungsmethode der Funktionalgleichungen von Typen $f(x+y)=\Sigma g_i(x)h_i(y)$.*

VINTI, CALOGERO and BAIADA, EMILIO, *Sur la validité de la propriété fondamentale de la transformée de Fourier.*

VOREL, ZDENEK, *Theory of Kirchhoff's networks from the viewpoint of ordinary differential equations.*

VRKOČ, IVO, Устойчивость при постоянно действующих возмущениях. (*Stability under constantly acting perturbations.*)

WIDOM, HAROLD, *Eigenvalues of n-dimensional convolution operators.*

WILF, HERBERT S., *A generalization of the inequality of arithmetic and geometric means and applications.*

WLOKA, JOSEPH, *Partielle, lineare Differentialgleichungen für vektorwertige Distributionen.*

YOOD, BERTRAM, *Non-commutative Banach algebras and almost periodic functions.*

YOUNG, LAURENCE C., *On the calculus of variations for parametric multiple integrals.*

ŻELAZKO, WIESŁAW, *Some remarks on the locally bounded algebras.*

ZIELEZNY, ZBIGNIEW, *Fourier transforms of distributions with one-sided bounded carriers.*

Section 4—Topology and Differential Geometry

ADLER, ALFRED W., *Classifying spaces for Kähler manifolds.*

VAN ALBADA, PIETER J., *On a generalization of a problem of H. Steinhaus.*

ANDERSON, L. W., and HUNTER, R. P., *Homomorphisms and dimension.*

ANDERSON, R. D., *Covering flows.*

ANDREATTA, ANTONIO, *A few remarks on the relative theory of graphs.*

ANGELITCH, T. P., *Eine Verallgemeinerung des Begriffs des Darboux'schen Vektors für den Raum von Riemann.*

ARHANGEL'SKIĬ, A. (Presented by P. S. ALEKSANDROV), О размерности пространств. (*On the dimension of spaces.*)

BARRATT, MICHAEL GEORGE, *The structure of primary homotopy operations.*

BARTHEL, WOLDEMAR, *Zur nichtlinearen Differentialgeometrie.*

BAUER, FRIEDRICH-WILHELM, *Ein Funktor der Homotopie in Homologie überführt.*

BERGAU, PETER, *Über reguläre Zerlegungen geschlossener Flächen (regular maps).*

BERSTEIN, I., *On cogroups in the categories of graded algebras and of c.s.s. groups and on principal cofibre bundles.*

BILINSKI, STANKO, *Eine Verallgemeinerung der Bertrandschen Kurven.*

BLANUŠA, DANILO, *Isometric imbeddings of Euclidean spaces in spherical spaces.*

BLUM, RICHARD, *On conformally Euclidean spaces of class one.*

BRACONNIER, JEAN, *Sur la convolution des courants et ses applications.*

CAIRNS, STEWART SCOTT, *Retractions by differentiable isotopies.*

CALAPSO, R., *Sulle congruenze di Ribancour.*

CATTANEO-GASPARINI, IDA, *Transverse covariant derivative and almost product spaces.*

CHURCH, P. T., *Differentiable open maps on n-manifolds.*

ČOGOŠVILI, G. S., Диаграммы групп с внешними умножениями и некоторые их применения в теории гомологии. (*Diagrams of groups with outer multiplication and some of their applications in homology theory.*)

COUTY, RAYMOND, *Transformations projectives des variétés presque Kählerienne.*

CSÁSZÁR, ÁKOS, *Complétion d'espaces non-séparés.*

CURTIS, MORTON L., and LASHOF, R. K., *On product and ball-bundle neighbourhoods.*

DANZER, LUDWIG, W., *Überdeckungen mit kongruenten Kugeln und Durchschnittseigenschaften von Kugelfamilien im E^n.*

DEMARR, RALPH, *Generalized homomorphisms in semigroups and multivalued mappings.*

DEMARR, RALPH, *Some results on common fixed points for commuting mappings.*

DIRAC, GABRIEL A., *Remarks on the four colour conjecture and the theory of graphs.*

DOITCHINOV, D. B., *Uniforme Einbettungen im Hilbert-Raum.*

ELIOPOULOS, HERMES A., *Connections of manifolds with almost tangent structure.*

EPSTEIN, DAVID B. A., *Steenrod operations in Abelian categories.*

EWALD, GÜNTER, *Konvexe Funktionen in der Theorie konvexer Körper.*

FADELL, EDWARD, *Whitney duality for locally flat topological manifolds.*

FÁRY, I., *The theory of Abelian group extension and an axiomatic of cohomology theory.*

GOETZ, ABRAHAM, *Special connections associated with a given linear connection.*

GOŁAB, STANISŁAW, *Algèbre des objets géométriques de Pencov.*

GOLDBERG, SAMUEL I., *The conformal transformation group of a Riemannian manifold.*

GREEN, LEON W., *Flows on solvmanifolds.*

HARROLD, ORVILLE G. JR., *Locally peripherally Euclidean spaces are locally Euclidean, II.*

HEMMINGSEN, ERIK, *Branch sets of light interior mappings.*

HSIUNG, CHUAN-CHIH, *Vector fields and infinitesimal transformations on Riemannian manifolds with boundary.*

HUNTER, R. P. and ANDERSON L. W., *On the H-equivalence in a semigroup.*

ILIĆ-DAJOVIĆ, M., *Deux théorèmes sur les déformations infiniment petites cylindroides.*

JUSSILA, OLLI K., *On a theorem of Floyd.*

KATSURADA, YOSHIE (Read by H. HOPF), *Some properties for closed hypersurfaces in Riemann space.*

KAWAGUCHI, AKITSUGU, *On invariants of higher order spaces.*

KEISLER, JAMES E., *Inverse incidence limits.*

KELDYŠ, L. V., Непрерывные разбиения Евклидова пространства. (*Continuous decompositions of Euclidean space.*)

KOSINSKI, ANTONI, *On the structure of combinatorial manifolds.*

KUREPA, DJURO, *On an inequality concerning cartesian multiplication.*

LAUDAL, OLAV ARNFINN, *Cohomologie locale. Applications.*

LEDERER, GEORGE, *A generalization of the notion of connectedness.*

LEICHTWEISS, KURT, *About curvature theory of manifolds in the Euclidean n-space.*

LIBERMANN, PAULETTE, *Connexions d'ordre supérieur.*

MCDOWELL, R. and DE GROOT, J., *Autohomeomorphism groups of 0-dimensional spaces.*

MANN, LARRY N., and SU, J. C., *Actions of elementary p-groups on manifolds.*

MARDEŠIĆ, SIBE, *ε-mappings onto polyhedra.*

MARKUS, LAWRENCE, AUSLANDER, L., and HAHN, F., *Minimal flows on nilmanifolds.*

MENNICKE. JENS, *Einige Ergebnisse zur Flächentopologie.*

MICHAEL, ERNEST A., *Trivial extensions.*

MOLNÁR, JÓZSEF, *Über Kreislagerungen.*

NÁDENÍK, ZBYNĚK, *Über die Enveloppe von konvexen Zylinderflächen.*

NITSCHE, JOHANNES C. C., *The extension of minimal surfaces intersected in starshaped curves by parallel planes.*

NOVAK, J., *The topological convergence groups.*

NOVIKOV, S. P. (Read by W. BROWDER), Гладкие многообразия общего гомотопического типа. (*Smooth manifolds of common homotopy type.*)

PAPIĆ, PAVLE, *Sur les images continues des compacts ordonnés.*

PASYNKOV, B. N. (Read by P. S. ALEKSANDROV), Об универсальных пространствах. (*On universal spaces.*)

PIER, JEAN-PAUL, *Sur les groupes topologiques compacts totalement discontinus.*

PIRANIAN, GEORGE, *The boundary elements of manifolds.*

POGORELOV, A. V., Об изометрическом погружении двумерного риманова многообразия в трехмерное риманово пространство. (*On the isometric immersion of a two-dimensional Riemannian manifold into a three-dimensional Riemannian space.*)

PONOMAREV, V. I. (Read by P. S. ALEKSANDROV), Проекционные спектры и топологические пространства. (*Projective spectra and topological spaces.*)
POST, KAREL, *Star extension of plane convex sets.*
REINHART, BRUCE L., *Periodic orbits of vector fields.*
RINGEL, GERHARD, *Selbstkomplementäre Graphen.*
RIZZA, G. B., *Finsler structures on almost complex manifolds.*
SABININ, L. V., Зеркальные симметрии однородных пространств и групп Ли. (*Mirror symmetries of homogeneous spaces and Lie groups.*)
SARANTOPOULOS, SPYRIDON, *La forme générale des équations des courbes de Bertrand.*
SCHERK, PETER, *Osculating spaces.*
SELFRIDGE, J. L., *On a problem in covering sequences.*
SHEPHARD, GEOFFREY COLIN, *Projection functions of a convex body.*
ŠNAJDER, Z., *Über eine natürliche Verallgemeinerung der Mongeschen Projektionsmethode auf mehrdimensionale Räume.*
STEER, BRIAN, *The Samelson product in F^n.*
STONE, ARTHUR HAROLD, *Classification of absolute Borel spaces.*
STRUBECKER, KARL, *Airysche Spannungsfunktion und isotrope Differentialgeometrie.*
THOM, RENÉ, *Equivalence topologique des applications polynomiales.*
TUTTE, WILLIAM T., *Enumeration of planar maps.*
VALETTE, GUY, *Les drapeaux et leurs éléments de contact.*
VINCENSINI, PAUL, *Surfaces harmoniques et représentation conforme.*
VRĂNCEANU, GHEORGHE, *Sur les espaces fermés à connexion affine.*
WARD, AUGUSTUS J., *The hyperspace of curves in a uniform space.*
WHITTAKER, JAMES *V.*, *Group isomorphisms of some homeomorphism groups.*
WILLE, R. J., *Rigid topological groups.*
WOLF, JOSEPH A., *Homogeneity of multiply connected Riemannian manifolds.*
WONG, YUNG-CHOW, *Existence of linear connections with respect to which a given tensor is parallel or recurrent.*
WRONA, WŁODZIMIERZ, *On a generalization of Schur theorem.*
YANO, KENTARO, *On a structure f satisfying $f^3+f=0$.*
ZEEMAN, ERIK C., *Topology of the brain.*

Section 5—Algebraic Geometry

CROWE, DONALD W., "*Distance*" *in a plane over* $GF(2^n)$.
DEMBOWSKI, PETER, *Partial planes with parallelism.*
ETAYO, JOSÉ-JAVIER, *The concept of algebraic equivalence of divisors of a field.*
GODEAUX, LUCIEN, *Surfaces algébriques régulières dépourvues de courbe canonique possédant un système bicanonique irréductible.*
GREENBERG, MARVIN J., *Realizations of schemata over local rings.*
HERRMANN, MANFRED, *Eine Multiplizitätsbestimmung mit Hilfe der Theorie der Nachbarpunkte.*
KARZEL, H. J., *Affine und projektive Gruppenräume.*
KEEDWELL, ANTHONY D., *A configurational proposition in the projective plane.*
MANIN, J. I., О классификации формальных абелевых групп. (*On the classification of formal Abelian groups.*)
MAYER, ALAN L., *A theorem of Riemann.*
OORT, FRANS, *Multiple algebraic curves.*

PEDOE, DANIEL, *A fundamental theorem in projective geometry.*
PORTEOUS, IAN ROBERTSON, *Thom polynomials in algebraic geometry.*
PETKANČIN, B., *Begründung der komplexen zweidimensionalen Möbiusschen Geometrie.*
REINHARDT, KIEHL, *Zur Reduktionstheorie algebraischer Mannigfaltigkeiten.*
SALZMANN, HELMUT, *Flat topological planes.*

Section 6—Probability and Statistics

BARTLETT, MAURICE STEVENSON, *Statistical estimation of density functions.*
BHARUCHA-REID, A. T., *Equivalent Markov processes.*
BIRNBAUM, Z. W., *Some simple distribution-free tests of fit.*
BODIOU, GEORGES, *Caractérisation du conditionnement quantique des lois de probabilité.*
BOL'ŠEV, L. N., Некоторые приложения пирсоновских преобразований. (*Some applications of Pearson transformations.*)
BRAUMANN, PEDRO B. T., *Eigenschaften der Normierungskonstanten von Gnedenko und Groshev.*
BRAUMANN, PEDRO B. T., *The multiplication of a finite number of not necessarily σ-finite measures.*
BRAUMANN, PEDRO B. T., *Bemerkenswerte Sonderfälle der schwachen Konvergenz von Summen unabhängiger Zufallsgrössen.*
BRAYTON, ROBERT K., *On the asymptotic behaviour of the number of trials necessary to complete a set with random selection.*
BRILLINGER, DAVID ROSS, *Tukey's general method of setting confidence limits (the Jack-knife) applied to the case of maximum likelihood estimates.*
COHEN, ALONZO CLIFFORD, JR., *Progressively censored samples.*
DANSKIN., JOHN M., *A game over spaces of probability distributions.*
DAVID, HERBERT A., *Designs for paired comparisons.*
EICKER, FRIEDHELM, *On consistent and asymptotically normal estimators for classes of linear regressions.*
FOGUEL, SHAUL R., *Contractions in Hilbert spaces and Markov processes.*
FURSTENBERG, HARRY, *A Poisson formula for semi-simple Lie groups.*
GAEDE, KARL-WALTER, *Anwendung einer stochastischen Anordnung bei Wartezeitproblemen.*
GANI, JOSEPH MARK, *On the age-distribution of replaceable ranked elements.*
GOOD, IRVING J., *A compromise between credibility and subjective probability.*
GURLAND, JOHN, *Estimation and testing hypotheses for the negative binomial and other distributions.*
HIORNS, ROBERT W., *Statistical estimation for linear compound models.*
HOEFFDING, WASSILY, *The strong law of large numbers for U statistics.*
IBRAGIMOV, I. A. (Presented by A. N. KOLMOGOROV), Об условиях слабой зависимости стационарных гауссовских процессов. (*On conditions for weak dependence for stationary Gaus- sian processes.*)
IGLEHART, DONALD L., *Competition processes.*
KAMPÉ DE FÉRIET, JOSEPH, *Sur la décomposition d'une fonction aléatoire normale.*
LAHA, R. G., *On an analytical decomposition of the Poisson law.*
LEIPNIK, R. B., GREEN, H. S., and MAXFIELD, J. E., *Algebras of lattice paths and the Ising problem.*
LEWIS, TOBIAS, *Quasicorrective errors.*

LINNIK, JU. V., О подобных зонах в математической статистике. (*On similar regions in mathematical statistics.*)

LUKACS, EUGENE, *A linear mapping of the space of distribution functions onto a set of bounded continuous functions.*

MASCHLER, MICHAEL, *Immune coalition-structures for n-person cooperative games.*

MULLER, MERVIN E., *Sequential (item by item) selection for non-sequential sampling plans.*

NABEYA, S., *On Pearson-Geary criteria in the analysis of variance.*

NARAYANA, T. V., *A combinatorial lemma and its application to probability and statistics.*

OREY, STEVEN, *Potential kernels for recurrent Markov chains.*

PALÁSTI, ILONA, *Threshold functions for subgraphs of given type of the bichromatic random graph.*

PATIL, GANAPATI P., *Best unbiased estimation for generalized power series distribution.*

PONTRJAGIN, L. S., Одна статистическая задача теории оптимального управления. (*A statistical problem in the theory of optimal control.*)

PRÉKOPA, A., *On stochastic programming.*

PYKE, RONALD, and HOBBY, CHARLES, *Combinatorial methods in the theory of sums of random variables.*

RÉNYI, ALFRÉD, *On the theory of outstanding observations.*

RIDER, PAUL R., *Incomplete moments of the binomial distribution.*

ROBBINS, HERBERT, and CHOW, Y. S., *A renewal theorem for non-identically distributed random variables.*

ROZANOV, JU. A. (Presented by A. N. KOLMOGOROV), Об эффективности метода наилучших квадратов для зависимых величин. (*On the effectiveness of the least squares method for dependent variables.*)

ROZANOV, JU. A. (Presented by A. N. KOLMOGOROV), К вопросу об эквивалентности вероятностных мер, отвечающих гауссовским стационарным процессам. (*On the question of equivalence of probabilistic measures corresponding to Gaussian stationary processes.*)

RUBIN, HERMAN, *Convergence in probability and convergence with probability one of nets.*

SAVAGE, I. RICHARD, *Partial orderings of probabilities of rank orders.*

SCHEFFER, CAREL LOUIS, *Bounds for the eigenvalues of Markov-matrices (continuous time).*

SCHNEIDER, BERTHOLD, *Die Verteilung von Schätzwerten bei nichtlinearer Regression.*

SIDDIQUI, MOHAMMED M., *Distribution of Student's t in samples from a rectangular universe.*

SNELL, J. LAURIE, *Boundary theory for recurrent Markov chains.*

SOBEL, MILTON, *On statistical search problems.*

STEPHENS, MICHAEL A., and WATSON, GEOFFREY S., *Tests for spherical normal and circular normal distributions.*

URBANIK, KAZIMIERZ, *Applications of information theory in quantum mechanics.*

VRANIČ, VLADIMIR, *On the use of duality in the theory of correlation.*

v. WALDENFELS, WILHELM, *Der infinitesimale erzeugende Operator eines stationären Markowprozesses.*

WAUGH, W. A. O'NEILL, *Growth rates for birth processes.*

WILKS, S. S., *Multivariate statistical outliers.*

ZAREMBA, STANISLAW K., *Estimators of the cumulants of the variables generating a linear process.*

Section 7—Applied Mathematics, Mathematical Physics and Numerical Analysis

ABLOW, CLARENCE M., *Initial motion of an elastic earth under blast loading.*

ACKROYD, RONALD T., *Application of functional analysis to the high speed numerical solution, with error bounds, of the Boltzmann equation for neutron transport.*

ADAMS, ERNST W., *A class of similar solutions for the velocity and the temperature boundary layer in planar and axially symmetric channel flow.*

ALTEN, HEINZ-WILHELM, *Schranken für die Nullstellen der Gegenbauerschen Polynome.*

ALTERMAN, ZIPORA, *Propagation of a seismic pulse in a sphere.*

ANOSOV, D. V. (Presented by JA. G. SINAĬ), Грубость геодезических потоков на компактных римановых многообразиях отрицательной кривизны. (*The roughness of geodesic currents in compact Riemannian manifolds of negative curvature.*)

AUMANN, ROBERT J., *Markets with a continuum of traders.*

BAREISS, ERWIN H., *On the spectral analysis of the periodic transport operator.*

BARNES, JOHN L., *Functional transforms in information theory.*

BASS, LUDVIK, *The motion of test charges in rapidly oscillating electromagnetic fields.*

BOCHENEK, KRYSTYN, *Some applications of asymptotic expansions theorems.*

BONDER, JULIAN, *Sur la structure tensorielle du système différentiel de la dynamique des gaz réels.*

BONFIGLIOLI, LUISA, *Nomographie anaglyphique.*

BÖRSCH-SUPAN, WOLFGANG, *A-posteriori error bounds for polynomial roots.*

BRAMBLE, J. H., and PAYNE, L. E., *Some inequalities for vector functions with applications in elasticity.*

BREMERMANN, HANS J., *Quantum-theoretical limitations of data processing.*

BROWNELL, FRANK H., *An exact computation model for the Lamb shift.*

BUSK, THØGER, *Derivatives computed by ordinary interpolation methods.*

CATTANEO, CARLO, *On Mach's principle.*

CESCO, R. P., *On the general solution of the problem of three bodies.*

CHAMBERS, LLEWELYN G., *Propagation in dispersive media.*

CLENSHAW, CHARLES W., *The application of a curve fitting program to the ship-fairing problem.*

COLEMAN, ALBERT JOHN, *Structure of the Fermion density matrix.*

COOKE, JOHN CAISTER, *Triple integral equations.*

DAVIS, A. M. J., *Completeness of 2-D eigenoscillations of water in a semicircular bowl.*

DENNIS, STANLEY C. R., *On the slow motion of a viscous fluid past cylindrical bodies.*

DESTOUCHES, JEAN-LOUIS, *Some mathematical problems connected with the functional theory of particles.*

DOLEŽAL, VÁCLAV, *Regularity of Kirchhoff's networks.*

DOU, ALBERTO, *Elastic behaviour of a longitudinally welded cylinder.*

DOUGLAS, JIM, JR., and GUNN, J. E., *A general formulation of alternating direction methods.*

DRESHER, MELVIN, *A zero-sum two-person game in sampling from a finite population.*

DROBOT, STEFAN, *On variational principles of classical fields.*

DUPARC, H. J. A., *On equidistant error detecting codes.*

EBERLEIN, WILLIAM F., *The spin model of space-time.*

ERGUN, A. N., *Self-superposable fluid motions.*

FADDEEV, L. D., Строение резольвенты оператора энергии системы трех частиц и задача рассеяния. (*Construction of the resolvent of the energy operator for a three particle system and the scattering problem.*)

FERRARI, ITALO, *Sur un théorème d'unicité pour les équations magnétodynamiques des fluides.*

FLEISHMAN, BERNARD A., *Superposition of synchronous periodic regimes in relay control systems.*

FORBAT, NICK H., *Détermination approchée de la solution périodique d'une équation différentielle hétéronome non linéaire.*

FREY, TAMÁS, *Abschätzungs- und Entwicklungsmethoden der Eigenwertaufgaben von Differentialgleichungssystemen.*

GAIER, DIETER, *Über die Diskretisierung der Gerschgorinschen Integralgleichung.*

GALMARINO, ALBERTO R. (Presented by L. DOLANSKY), *On the cardinality of solutions of four-person constant-sum games.*

GAUTSCHI, WALTER, *Construction of Gaussian quadrature formulas.*

GHAFFARI, ABOLGHASSEM, *On an extension of the stroboscopic method.*

GHOSH, N. L., *Circulation lift and drag on aerofoils.*

GILBERT, EDGAR N. (Presented by F. W. SINDEN), *Models of polycrystalline solids.*

GILLIS, JOSEPH, and WEISS, GEORGE, *Viscous flow between non-parallel plane walls.*

GNANADOSS, ADAIKALAM ALBERT, *Fechner quantum and equal temperament.*

GOLDENBERG, HARRY, *Derivation of three-dimensional from two-dimensional solutions of transient heat conduction problems.*

GREENSPAN, HARVEY P., *The stability of a contracting shear layer.*

HANDELMAN, GEORGE H., and KELLER, JOSEPH B., *A singular perturbation method for eigenvalue problems.*

HASELGROVE, COLIN B., *Non-linear ordinary differential equations with two-point bondary conditions.*

HAYES, JAMES G., *A curve-fitting problem with six independent variables.*

HELLIWELL, JOHN B., *Gas-ionising shock and detonation waves in magnetogasdynamics.*

HERRIOT, JOHN G., *Calculation of particular solutions of linear partial differential equations by the method of integral operators.*

HOLT, MAURICE, *Calculation of nozzle flows by the method of integral relations.*

HUAUX, AIMÉ F. P. E., *Sur l'existence de solutions périodiques de certaines classes d'équations différentielles.*

ISAACSON, EUGENE, *An aspherical membrane under internal pressure.*

ITOH, MAKOTO, *The m-valued δ-algebra and its application to the theory of automata.*

JACOB, C., *Sur quelques problèmes aux limites de la mécanique des milieux continus.*

JANKOVIĆ, ZLATKO, *Some remarks on the tensor calculus.*

KARUSH, WILLIAM, and BELLMAN, RICHARD E., *On the maximum transform.*

KASTELEYN, P. W., *Arrangements of dimers on lattice graphs.*

KELLER, H. B., KELLER, J. B., and REISS, E. L., *Buckled states of circular plates.*

KLEIN, JOSEPH, *Notion de tenseur force en mécanique classique et relativiste.*

KOOPMANS, TJALLING C., and WILLIAMSON, RICHARD, *Utility and time: An axiomatic discussion.*

KRISTENSEN, POVL, *On the mathematical formulation of quantum field theory.*

KUHN, HAROLD W., *On the Pauli exclusion principle with 2-body interactions.*

KUSTAANHEIMO, PAUL E., *Über die Spinorringalgebra.*

LAITONE, EDMUND V., *Spherical surface harmonic analysis of numerical data.*

LAUWERIER, H. A., *A note on Sommerfeld's problem of diffraction by a semi-infinite screen.*

LAVRENT'EV, M. A., Некоторые краевые задачи теории потенциала. (*Some boundary problems in potential theory.*)

LAVRENT'EV, M. M., Об операторных уравнениях I-го типа и некоторых задачах теории потенциала. (*On operator equations of the first kind and certain problems of potential theory.*)

LUNDQUIST, STIG, *On a stability problem in reactor theory.*

MCCONNELL, JAMES, *Rotations and angular momenta.*

MCCREA, WILLIAM HUNTER, *A cosmological application of information theory.*

MAGIROS, DEMETRIOS G., *An actual solution of a problem of nonlinear mechanics.*

MARTIN, J. L., *Point interactions and the Lee model.*

MICHEL, J. G. L., *Numerical integration of ordinary differential equations through certain types of singularities.*

MORAWETZ, CATHLEEN S., *Approach to steady state for the wave equation.*

MOREAU, JEAN J., *Sur la naissance de la cavitation dans un fluide incompressible.*

NAKASHIMA, KATSUYA, *On the computation of eigenvalues of general matrices.*

NIEMEYER, H. F., *Über Eingenschwingungen elastischer Körper.*

ORLOFF, CONSTANTIN, *La résolution des équations différentielles au moyen des spectres mathématiques.*

PECK, JOHN E. L., *Polynomial approximation of observational data subject to constraint.*

PERRY, CLAY L., and ABLOW, C. M., *Numerical solution of the Dirichlet problem for the quasi-linear elliptic equation* $\Delta u = buu$.

PHAM, MAU QUAN, *Dynamique analytique relativiste du point.*

PILLOW, ALBERT F., *Forced convection of a viscous fluid at low Reynolds number past a circulation producing dipole in two-dimensions.*

REZA, F. M., *Generating theorems for positive real functions.*

ROSS, RODERICK A., *The fundamental solution of the thermoelastic wave equation.*

SALZER, HERBERT E., *Equally-weighted formulas for numerical differentiation.*

SARAFYAN, DIRAN, *A unified and general method of numerical solution of ordinary differential equations.*

SCHECHTER, SAMUEL, *Numerical solution of discrete variational problems.*

SETH, BHOJ RAJ, *Turning point analysis of plastic deformation.*

SHERMAN, SEYMOUR, *The Ising model for ferromagnetism.*

SINDEN, FRANK W., *A canonical form for quadratic programs.*

SOPKA, JOHN J., *Some abstract aspects of numerical quadrature.*

STANDISH, CHARLES J., *Remarks on the sampling theorem.*
SYNGE, JOHN L., *The Hamiltonian method applied to water waves.*
TAMMI, OLLI, *On analytic foundations of central projection.*
TAUB, ABRAHAM H., *Consequences of variational principles in general relativity.*
TRAUB, JOE FRED, *On the informational efficiency of iteration functions.*
TUCKER, ALBERT W., *A linear-convex existence theorem.*
VARGA, RICHARD S., *Higher order stable implicit methods for hyperbolic partial differential equations.*
WACHSPRESS, EUGENE L., *Optimum parameters for alternating-direction- implicit iteration.*
WASEL, ALBERT D., *On a variational theorem.*
WAŻEWSKI, TADEUSZ, *Équations au contingent et systèmes de commande.*
WENDROFF, BURTON, and LAX, PETER, *On the stability of difference schemes.*
WERNER, PETER, *An interface problem in the theory of acoustic wave fields.*
WIGHTMAN, A., *Some results on relativistic quantum field theory.*
WILCOX, CALVIN H., *Diffraction in inhomogeneous anisotropic media.*
WILKINSON, JAMES H., *Error bounds for unitary equivalence and similarity transformations.*
WILLIAMS, DAVID E., *A solution of the aerofoil integral equation by the Hilbert–Schmidt method and some related topics.*
WINOGRADZKI, JUDITH, *Formalisme tensoriel et spinoriel incluant les parités.*
ZIN, GIOVANNI, *General theory of the Čerenkov radiation.*
ZLÁMAL, MILOŠ, *Error estimates for positive definite boundary-value problems.*

Section 8—Education

ACTON, FORMAN S., *The growing importance of mathematical models in medical research.*
BRIDGER, MARCUS, *The use of desk calculating machines in schools.*
BUNT, LUCAS N. H., *Statistics in schools; basic notions for testing a hypothesis.*
FARAGÓ, L., *Difficulties of the analytical operation of thinking in the learning of mathematics.*
FEHR, HOWARD F., *Instruction in geometry for the secondary school.*
HARKIN, D. C., *Mathematics and language.*
HERRIOT, SARAH T., *Impact of the new school mathematics study group curriculum on college-bound students.*
HOOD, RODNEY T., *Inverting the conics.*
ILIĆ-DAJOVIĆ, M., *Sur la réalisation d'un enseignement moderne de géométrie.*
MASCHLER, MICHAEL, *Mathematics curriculum for humanistic studies.*
MAY, KENNETH O., *Undergraduate research in the United States.*
NESS, WILHELM, *Über die harmonische Reihe.*
ORLOFF, CONSTANTIN, *Concrétisation des notions algébriques dans l'enseignement secondaire.*
PERKUS, R., *What mathematics for the twelfth year?*
SAWYER, WALTER WARWICK, *Not modern; not traditional; but mathematics as a whole.*
SHANKS, MERRILL E., *The role of rigor in school mathematics.*
SMITH, E. M. R., *The challenge and opportunities of mathematics teaching in Africa.*
SUPPES, PATRICK, *The learning of mathematical concepts.*
THWAITES, BRYAN, *An experiment in new syllabuses.*
WIGAND, KLAUS, *Modernisierung des Mathematikunterrichts durch mathematische Geräte und praktische Verfahren.*

SECRETARY'S REPORT

PREPARATIONS

At a session on 28 October 1958 the Swedish National Committee for Mathematics decided to accept the invitation, conveyed by the International Mathematical Union, to organize the next International Congress of Mathematicians in Stockholm in 1962. This decision was endorsed by the Swedish Mathematical Society on 30 November 1958 and a joint invitation was issued to the mathematicians of the world, signed by the chairmen of the National Committee and the Society, Professors Å. Pleijel and G. Borg.

Professors O. Frostman and Å. Pleijel were entrusted with the initial preparations for the congress together with Professor H. Cramér, at that time Chancellor of the Universities of Sweden, and Professor L. Gårding, all of whom later constituted the Organizing Committee.

The scientific programme was drawn up in close cooperation with the International Mathematical Union, which for this purpose nominated a Consultative Committee with Professor G. de Rham as chairman. The Swedish representatives in charge of the scientific programme were Professors L. Gårding, L. Carleson and L. Hörmander.

Thus a first meeting was held in Zürich in November 1960 followed by a meeting in Düsseldorf in January 1961. As a result of these two meetings the "International Fields Committee", that elects the two prize winners, was constituted. Furthermore a list was made of speakers to be invited to deliver one-hour addresses on chosen topics in different fields of mathematics. Finally chairmen of international panels were appointed with the task of recommending half-hour speakers who would be able to present recent results within their respective fields. After receipt of the reports from the panels a final meeting was held in Zürich in November 1961; the choice of speakers to be invited to deliver half-hour lectures was then made.

About 1 June 1961 invitations to hold one-hour lectures at the Congress were sent to 19 mathematicians.

At the end of August 1961 a "First Communication" requesting those interested to notify the Secretariat of their wish to receive further information was distributed to all mathematicians listed in the World Directory of Mathematicians, 1958 edition, as well as to national committees, universities, academies, societies, etc. A supplementary distribution was made when in September the new edition of the World Directory appeared.

Already in 1960 the Organizing Committee had appointed a lawyer, Mr. R. Thörn, from the State Railway Board, to be the Secretary General of the Congress. From 1 September 1961 a permanent secretariat was set up at the Mittag-Leffler Institute in Djursholm.

About 1 December 1961 invitations to give half-hour lectures were sent out to 69 mathematicians, a few additional invitations being issued later.

At about the same time the "Second Communication" with Application Forms, Abstract Forms for short communications and other printed material was distributed.

Invitations to appoint delegates were also sent to the institutions referred to above.

Finally the "Third Communication", with a list of speakers who had already accepted the invitation, as well as other details about the Congress, was distributed in May 1962 to all those who had returned their Application Form.

FINANCE

Before accepting the invitation from the International Mathematical Union to organize the 1962 Congress, the Swedish National Committee for Mathematics had obtained assurance of financial support from the Swedish Government and from the Union. Support was subsequently obtained from a number of public and private bodies as listed above.

The membership fee was 160 Sw. Kronor for Ordinary, and 80 Sw. Kronor for Associate Members. For administrative reasons it was desirable to receive applications for membership as early as possible. Higher fees, 175 and 100 Sw. Kronor respectively, were therefore applied for payments after 28 February 1962. The fees were considerably higher than for the previous congresses, but in fact, quite normal in Sweden for a congress of this type and size.

THE CONGRESS

The Congress was preceded by the General Assembly of the International Mathematical Union on 11–13 August in Saltsjöbaden near Stockholm. The Secretariat of the Congress had already moved from Djursholm to the Royal Institute of Technology in Stockholm where the Congress Headquarters were set up and where most of the lectures were held. The Headquarters opened for registration at 10.00 on 14 August. Officially the Congress began with the Opening Ceremony which took place in the Konserthuset (Concert Hall) on Wednesday 15 August at 10.00 in the presence of His Majesty the King of Sweden, who had graciously consented to be patron of the Congress, and to present the Fields Medals. The closing session was held in the same hall on Wednesday, 22 August (special reports of these sessions are given below).

Besides the lecture rooms in the Royal Institute of Technology, two large halls with technical equipment were used for one-hour and half-hour lectures.

Of those attending the Congress, 2107 were ordinary members, and 984 were associate members.

These figures both separately and combined, exceed those for any of the previous congresses of mathematicians, as well as for any other scientific congress held in Sweden.

SCIENTIFIC PROGRAMME

The final scientific programme included 16 one-hour addresses of general interest; 2 of these were not given by the invited speakers in person.

The invited half-hour lectures as well as the short (ten-minute) communications offered by members were divided into the following sections:

1. Logic, Foundations and History
2. Algebra and Theory of Numbers

3. Analysis
4. Topology and Differential Geometry
5. Algebraic Geometry
6. Probability and Statistics
7. Applied Mathematics, Mathematical Physics and Numerical Analysis
8. Education

The number of half-hour lectures announced in the programme was 60; of these 55 were held by the invited speakers in person, 2 were read by deputies and 3 were cancelled.

Mimeographed booklets containing outlines of one-hour and half-hour addresses as well as translations of Russian texts were distributed to ordinary members on arrival. At the same time printed abstracts of short communications were also distributed. The number of such ten-minute communications given at the Congress was 745. The lectures were presided over by chairmen invited by the Organizing Committee.

Within section 8 there were also three special meetings organized by the International Commission on Mathematical Instruction with reports followed by discusssion on the following topics:

1. Which subjects in modern mathematics and which applications of modern mathematics can find a place in programmes of secondary school instruction?
   (Reported by J. G. Kemeny.)
2. Connections between arithmetic and algebra in the mathematical instruction of children up to the age of 15.
   (Reported by S. Straszewicz.)
3. Education of the teachers for the various levels of mathematical instruction.
   (Reported by K. Piene.)

In addition to the announced programme Professor S. L. Sobolev gave a special report on the use of electronic computers for the deciphering of the Maya language made at the University of Novosibirsk.

SOCIAL PROGRAMME

A social evening with buffet and dancing was held in the City Hall of Stockholm on Thursday, 16 August.

On the evening of Friday, 17 August, another social gathering with a song recital by the Swedish tenor, Nicolai Gedda, and a programme of folk dancing was arranged at Skansen, the open-air museum in the Djurgården in Stockholm.

Sunday, 19 August, was devoted to various whole-day excursions. Some excursions were made by steamer to the outer and inner Archipelago of Stockholm, the destinations being Sandhamn, Utö and Vaxholm respectively. There was also an excursion by motor coach to the Castle of Gripsholm, situated west of Stockholm on the shore of Lake Mälar, to the University City of Uppsala and to Sigtuna, capital of Sweden in the 12th century.

The evening programme for Tuesday, 21 August, included three alterna-

tive entertainments. A performance of Swedish ballet took place in the Royal Opera House. Two performances of Pergolesi's opera "Il maestro di musica" were given at the 18th Century Court Theatre of Drottningholm (an additional performance had been arranged on Monday, 20 August). Finally there was a concert of chamber music in the small auditorium of the Concert Hall.

On four mornings associate members had the opportunity to participate in excursions, sightseeing trips, visits to museums, exhibitions, institutions and other places of interest.

OTHER ARRANGEMENTS

The Stockholm office of Thos. Cook & Son had been appointed to provide accomodation for members and to make the technical arrangements for the Sunday excursions as well as the morning excursions for associate members. For the completion of these tasks the office had set up special Accomodation and Travel Bureaux within the Headquarters building. The Svenska Handelsbanken, the Post Office and the Telegraph Office had established temporary branch offices.

The Organizing Committee provided members with a ticket entitling them to free travel on all local trams, buses and underground trains within Stockholm during the Congress.

EXHIBITIONS

Two exhibitions of books were located within the Headquarters building, a general one arranged by Almqvist & Wiksell Bokhandel AB in collaboration with publishers from many countries and a special one of school text books organized by the International Commission on Mathematical Instruction. Exhibitions dealing with computers and other technical equipment of particular value for members interested in numerical methods and applied mathematics, were arranged by the International Business Machines Svenska AB, the Facit Electronics AB and the Telefon AB L. M. Ericsson.

OPENING SESSION

The opening session of the Congress was held in the Concert Hall on Wednesday, 15 August 1962, at 10.00 in the presence of His Majesty the King of Sweden.

An orchestra from the Royal Navy played a selection of Swedish music.

Professor O. Frostman, Chairman of the Organizing Committee, welcomed the audience with the following words:

Your Majesty, Ladies and Gentlemen:

On behalf of the Swedish mathematicians I wish you welcome to the International Congress in Stockholm, the Opening Session of which is honoured by the presence of His Majesty the King of Sweden. We are very glad that so many from all parts of the world have responded to our invitation and we hope that the addresses given and the new results presented here will be of great value for the development of our science. We also hope that the personal contacts, renewed and established at the Congress, will serve the same purpose and, at the same time, be links in the chain that unites us all.

At the International Congress in Cambridge, England, in 1912, Professor G. Mittag-Leffler invited the mathematicians to meet the next time in Stockholm in 1916. Because of the First World War this Congress had to be reduced to a Scandinavian one, and it was not until after the Edinburgh Congress in 1958 that the Swedish mathematicians assumed the responsibility of organizing the next Congress. It goes without saying that the decision was a hard one to take. In the first place the number of participants at an international congress of mathematicians would probably have risen to something between 2000 and 3000 by 1962—in fact, we are over 3000, associated members included—and such figures must inevitably involve difficulties of housing, transportation, too small lecture-rooms, and so on. But the real trouble lies in the development of mathematics itself which is proceeding so rapidly that no man can survey but a part or parts of the front line, and total coverage can only be achieved by joint work on an international basis.

To be able to present a scientific programme worthy of an international congress it was therefore decided at an early stage to seek the assistance of the International Mathematical Union, and at a meeting in Zurich in November 1960 a small Consultative Committee was appointed with Professor de Rham, Lausanne, as chairman. The wide experience and knowledge represented in the Consultative Committee itself and strengthened by contacts with experts from all over the world, made it possible to choose the subjects and speakers for the one-hour addresses and to appoint chairmen of the international panels which have proposed the half-hour speakers. At subsequent meetings the Consultative Committee brought the information gathered to the Swedish representatives and all decisions were made in agreement. It must be clearly stated that the Swedish Committee takes the full responsibility for the organization of the congress, but without the invaluable help of the panels and the Consultative Committee the scientific programme would not have been adequate.

The part performed by the International Mathematical Union in preparing the scientific programme of this congress is a leading one, and is well suited to

act as a precedent for any future international congress. It seems therefore quite natural that the President of the International Mathematical Union should preside over the general sessions of the Congress, and I now have the honour to call upon Professor Rolf Nevanlinna who will declare the Congress open.

Professor R. Nevanlinna came forward and spoke as follows:

Your Majesty, Ladies and Gentlemen:

One prime characteristic of our age is the highly accelerated pace of its development. This can be felt everywhere in our cultural, social and economic life. It applies with greatest force to technology and science.

Mathematics is no exception. To convince oneself of this, it is sufficient to glance at the questions dealt with in the lectures and discussions of the international mathematical congresses since the beginning of the century.

The expansion and growth of many different branches of science have led to increasing specialization in the field of research. Today there is no mathematician who can claim to have mastered modern mathematics in its entirety.

This development of science would soon lead to an impossible situation, if there were not another tendency working against it; the tendency towards synthesis. Mathematical development in our century stands out as a shining example of this synthesis, which is necessarily on a conceptually high and abstract level.

What takes place in pure mathematics has its roots in the world of experience. On the other hand, theoretical and general mathematical insight throws light on practical questions, and forms, in fact, the basis of applications in many different fields: in natural science, in technology and, in recent times, in many branches of social and economic life. The astonishing development of electronic computers has contributed enormously to the applicability of mathematical methods.

Thus, mathematics in our time forms a background of ever increasing importance for all cultural life.

Mathematicians from all over the world have come together again after four years to survey the state of our science.

This survey is the principal object of a large Congress, and this opinion has been decisive in the organization of this Congress.

The most important branches of mathematics will be covered in 16 one-hour addresses, delivered by leading experts. These addresses are meant primarily for those participants of the Congress who are not specialists in these fields.

It is my hope that these survey-lectures will contribute towards a better contact between the different branches of mathematics.

About 60 half-hour addresses by invited speakers in the seven sections of the congress will report on the latest advances in their respective fields of work.

Freedom of research requires that all mathematicians of the world, and in particular those of the younger generation, should have the opportunity to present briefly their own results during this Congress. For practical reasons, however, the duration of such short communications, of which there are about 800, must be restricted to 10 minutes each.

So I hope that the week beginning today will supply proof that large congresses are still useful and have a task to fulfil as supplements to the smaller meetings at which a small group of specialists discusses their problems.

We are grateful to our Swedish colleagues for having taken over the great

burden of organizing this Congress and having given us the opportunity of meeting in this beautiful city of Stockholm.

In the name of the International Mathematical Union and of the Swedish Organizing Committee I have the honour to declare open the International Congress of Mathematicians in Stockholm.

After an interval, Professor Nevanlinna in his capacity of chairman of the Fields Medals Committee, read the report of the Committee which was as follows:

Your Majesty, Ladies and Gentlemen:

During the International Congress of Mathematicians in Zürich 1932, it was made known that J. C. Fields, chairman of the Organizing Committee of the Toronto Congress in 1928 had set up a trust to found two prizes to be presented in connection with the International Congress of Mathematicians that followed. This idea was carried out for the first time at the Congress in Oslo, 1936. At this and at the following Congresses in Cambridge (Harward) 1950, Amsterdam 1954 and Edinburgh 1958, two gold medals and cash prizes were presented to two young mathematicians in recognition of distinguished achievements in mathematics.

The problem of suggesting names for the award of the Fields medals has, since the last congress, been entrusted to the International Mathematical Union. To prepare the names at this Congress in Stockholm, the Union appointed a Fields medals committee. This committee has the honour to make known its decision here.

The two Fields medals are to be given, also this time, to two young mathematicians for distinguished scientific achievements: LARS HÖRMANDER, Professor at the University of Stockholm, and JOHN MILNOR, Professor at the University of Princeton.

The International Mathematical Union considers it a great honour that His Majesty the King has agreed to be present here and to give the Fields Medals to the winners of the Prizes.

May I now request Professor Hörmander and Professor Milnor to come forward to receive the Prizes from the hands of His Majesty.

His Majesty the King then presented the awards to Professors L. Hörmander and J. Milnor and complimented the prize-winners on their work.

The formal opening ceremony had thus come to an end and His Majesty withdrew from the session.

After a short interval reports on the work of the Fields Medallists were given. Professor L. Gårding spoke on the work of Hörmander and Professor H. Whitney on that of Milnor. The addresses are included in these Proceedings.

CLOSING SESSION

The Closing Session was also held in the Concert Hall, on Wednesday, 22 August 1962, at 14.00.

Professor Nevanlinna opened the session with the following words:

Ladies and Gentlemen

At every international congress of mathematicians since 1950, it has been the custom, on the last day of the Congress, to make an announcement about the International Mathematical Union. Following that custom, I am glad to announce that the Fourth General Assembly of the International Mathematical Union which met in Saltsjöbaden from August 11 to 13 was very successful. The Union, as some of you may know, has as one of its principal objectives the support of the International Congress of Mathematicians. I am glad to say that as a result of the Union's efforts, there has evolved, and become established, a definite pattern of collaboration with the Congress, especially with regard to the Fields Medals, the Scientific Program, and the choice of location of future Congresses. This, in itself, is by no means a trivial development, since the Union, as the only non-governmental organization, truly scientific in its intent, and truly international in its scope, offers perhaps the best guarantee of enduring international collaboration among mathematicians the world over.

Professor Nevanlinna then reported that the Executive Committee of the International Mathematical Union for the years 1963–66 would consist of Professors G. de Rham, President, H. Cartan and K. Kuratowski, Vice Presidents, K. Chandrasekharan, Secretary, and J. C. Burkill, F. Hirzebruch, M. A. Lavrentiev, D. Montgomery, B. Segre, and R. Nevanlinna (Past President), Members.

He also asked the meeting to join him in wishing the new Executive Committee every success, which was carried by acclamation.

Professor Nevanlinna continued:

Ladies and Gentlemen:

As Chairman of the Committee to recommend the choice of location of the International Congress of Mathematicians 1966, I have the honour to request Academician Lavrentiev to address this gathering.

M. Lavrentiev, j'ai l'honneur de vous donner la parole.

Professor M. A. Lavrentiev came forward and read an invitation—to hold the next Congress of Mathematicians in the Soviet Union—worded as follows:

Господин президент, дамы, господа!

От имени Академии Наук Советского Союза имею честь предложить конгрессу созвать следующий международный математический конгресс в Советском Союзе. За последнее десятилетие мировая математика достигла замечательных успехов и здесь присутствуют многие, усилиями которых были обеспечены эти успехи. Интересно то, что получили богатые приложения из познания техники, даже те разделы математики, которые раньше

казались отвлеченными и далекими от жизни. Господин президент, дамы, господа! Бурный рост математики, увеличение числа ее направлений и работников в них, все это увеличивает роль международной математической организации, роль объединения усилий математики всех стран на решение важнейших проблем. Одной из важных форм международного общения являются математические съезды. Разрешите от имени советских математиков здесь заявить, что мы приложим все усилия для того, чтобы совместно с исполнительным комитетом международного союза подготовить и провести новый съезд с наибольшей пользой и удовольствием для членов съезда и их семей.

The Russian text of the invitation was followed by an English interpretation given by Professor P. S. Aleksandrov.

The invitation was carried by acclamation, interpreted by Professor Nevanlinna with the following words:

M. Lavrentiev, le Congrès accepte votre invitation à l'unanimité. Je vous remercie sincèrement.

Professor S. Eilenberg asked for permission to speak and addressed the meeting as follows:

Mr. President, Ladies and Gentlemen:

In the name of the mathematicians of all nations assembled here, I take the liberty to say a few words which I hope represent the feelings of all of us here.

First, it is my pleasant duty to ask the Organizing Committee to convey expressions of our thanks and gratitude to His Majesty the King of Sweden for having placed this Congress under His high patronage and for having graced the opening session with His presence.

The physical arrangements for this congress were entirely in the capable hands of the Organizing Committee composed of Swedish mathematicians. The Swedish mathematical community being relatively small as compared with this huge congress, their task was enormous and the accomplishment most remarkable. I hope that I speak for all of us when I express our sincere thanks to the Organizing Committee for their hard task so well carried out.

In preparing the scientific programme, the Organizing Committee worked in close cooperation with the Executive Committee of the International Mathematical Union, and enlisted the help of many advisory committees composed of experts of all nations. This truly international effort resulted in a scientific programme of the highest caliber. Let me express here the hope that this procedure constitutes a precedent for future International congresses.

Having presented his personal thanks in Swedish to the Organizing Committee and especially to its Chairman, Professor Frostman, Professor Nevanlinna terminated the session with the following words:

Ladies and Gentlemen:

All good things must come to an end; so must this Congress. It has been a very successful meeting. We have witnessed many brilliant contributions to our science. We have had the pleasure of meeting many eminent personalities. But the dominant characteristic of this Congress has been the vigorous role played by youth. It seems to me that this is the best symbol of its success.

Now is the time to give our thanks to our Swedish Colleagues who made this

possible. They have worked hard, and long, and unselfishly; they have been modest and un-obtrusive. But, you will agree, they have been kind, and generous and efficient. The Organizing Committee, and especially its chairman Professor Frostman, have placed us in debt. Let us discharge that debt by expressing, publicly, our deep gratitude.

Ladies and Gentlemen, I declare the International Congress of Mathematicians in Stockholm closed.

In compliance with the demand expressed by Professor Eilenberg, the Organizing Committee sent the following telegramme to His Majesty the King:

On behalf of the mathematicians assembled in Stockholm for the International Congress the Organizing Committee thanks Your Majesty most warmly both for becoming patron of the Congress and for graciously attending the opening session and presenting the Fields medals.

THE WORK OF THE FIELDS MEDALLISTS

HÖRMANDER'S WORK ON LINEAR DIFFERENTIAL OPERATORS

By L. GÅRDING

Lars Hörmander was born in 1931, studied at Lund University and is now professor of mathematics at Stockholm University. He has been given a Fields Medal for his outstanding work in the theory of partial differential equations. Before trying to describe some of his results I shall have to lead you through a somewhat lengthy introduction.

Let me first acquaint you with a few notations. Coordinates and derivatives in real n-space will be denoted by x_k and $D_k = i^{-1}\partial/\partial x_k$ respectively. Higher derivatives will be written as powers $D^\alpha = D_1^{\alpha_1} \dots D_n^{\alpha_n}$; we put $|\alpha| = \alpha_1 + \dots + \alpha_n$. A linear differential operator of order m is a sum

$$P = P(x, D) = \sum_{|\alpha| \leqslant m} a_\alpha(x) D^\alpha$$

with principal part

$$P_m(x, D) = \sum_{|\alpha| = m}$$

We assume that the coefficients are smooth complex functions. Substituting complex numbers ξ_j for the derivatives we get the characteristic polynomials

$$P(x, \xi) = \sum_{|\alpha| \leqslant m} a_\alpha(x) \xi^\alpha$$

and $P_m(x, \xi)$. The Fourier transform will convert a differential operator with constant coefficients $P(D)$ into multiplication by $P(\xi)$. The adjoint and the conjugate of P are defined by

$$P^* = \sum D^\alpha \bar{a}_\alpha(x)$$

and

$$\bar{P} = \sum \bar{a}_\alpha(x) D^\alpha.$$

Linear differential equations

$$Pu = v \tag{1}$$

and systems of such equations occur in many mathematical models of physics, e.g. hydrodynamics, elasticity, thermodynamics, electricity and magnetism and magnetohydrodynamics. We can think of u as a function describing some aspect of the state of the physical system and v as some kind of exterior force. The manifold of solutions u gives us the possible states under the action v. Additional restraints mostly given as boundary

conditions give us a solution describing a particular situation. In practice it is the boundary conditions that account for the variety while P is given by the model itself.

It was pointed out very emphatically by Hadamard that it is not natural to consider only analytic solutions and source functions even if P has analytic coefficients. This reduces the interest of the Cauchy-Kowalevski theorem which says that (1) has locally analytic solutions if P and v are analytic. The Cauchy-Kowalevski theorem does not distinguish between classes of differential operators which have, in fact, very different properties.

If

$$P = D_1^2 + \ldots + D_n^2$$

is Laplace's operator, we are dealing with potential theory and if

$$P = D_1^2 - D_2^2 - \ldots - D_n^2$$

is the wave operator, we are studying wave propagation. The solutions of the homogeneous equation $Pu = 0$ behave quite differently. In the first case they are harmonic and hence real analytic, in the second case the example $u =$ arbitrary function of $x_1 \pm x_2$ shows that they may have very complicated singularities.

A natural reaction to all this is the question: what is it in P that makes this difference? More generally we can ask for properties of P or rather its characteristic polynomial which are intrinsic in the sense that they are more or less equivalent to properties of the solutions. Questions of this nature have no physical background but a very solid motivation: mathematical curiosity. They lead Hadamard to the fruitful notion of correctly set boundary problems. The first complete results, however, are due to Petrovsky who proved among other things that

$$Pu = 0 \Rightarrow u \text{ analytic} \tag{2}$$

and

$$\xi \text{ real} \neq 0 \Rightarrow P_m(\xi) \neq 0 \tag{3}$$

are equivalent properties for operators with constant coefficients. The second condition has an immediate extension to operators with variable coefficients. They are said to be elliptic. A simple example is Laplace's operator. Actually Petrovsky's main achievement on that occasion is a proof that all solutions of suitably defined elliptic non-linear analytic systems are analytic. The problem goes back to one of Hilbert's problems from the Paris congress in 1900.

Petrovsky felt that his results were just a beginning and in a lecture in 1945 he explicitly asked for a general theory of linear differential operators including those which do not appear in the mathematical models of physics. At the same time the theory of distributions appeared as a new tool in analysis. In his book on distributions Laurent Schwartz pointed out that the equivalence of (2) and (3) holds if the first u in (2) is assumed to be a distribution. He also stated a number of problems about differential operators. Since then a rather comprehensive theory has been worked out. Many people have contributed but the deepest and most significant results are due to Hörmander. I can describe only a few of them.

One of Schwartz's questions was the following one: what becomes of (3) if we replace (2) by the weaker statement

$$Pu=0 \Rightarrow u \in C^\infty ? \tag{4}$$

The answer, namely

$$\text{Im } \xi \text{ bounded}, \quad \xi \to \infty \Rightarrow P(\xi) \to \infty,$$

was obtained by Hörmander in his thesis as a byproduct of a study of operators with constant coefficients, which I shall not go into. The corresponding operators are called hypoelliptic. A characteristic point of the proof is that the theorem of the closed graph is used to replace (4) by an equivalent inequality which establishes the problem in a suitable analytical form. The last part of the thesis deals with variable coefficients and establishes an important inequality which can be described as follows.

Let Ω be an open part of R^n and let H_0^k be the closure of $C_0 = C_0^\infty(\Omega)$ in the norm

$$|u|_k^2 = \int \sum_{|\alpha| \leqslant k} |D^\alpha u(x)|^2 dx.$$

It is a Hilbert space and we can identify its dual H^{-k} with the space of distributions u in Ω for which

$$|u|_{-k} = \sup |(u,v)| / |v|_k < \infty,$$

where $(u,v) = \int u(x)\overline{v(x)}dx$ and v runs over C_0. We observe that $H^{-0} = H_0^0$ and that H^{-k} increases with k. A simple result from the theory of linear operators tells us that if A and B are Banach spaces and L is a densely defined linear operator such that

$$A \xrightarrow{L} B \quad \text{has a continuous inverse}, \tag{5}$$

then the adjoint map

$$B^* \xrightarrow{L^*} A^*$$

is onto. By a piece of ingenious and powerful analysis Hörmander established (5) in the form of an inequality

$$|u|_{m-1} \leqslant C|Pu|_0 \quad (u \in C_0)$$

and hence also the solvability of

$$P^*u = v \in H^{1-m}. \tag{6}$$

It is assumed that

$$\Omega \text{ is sufficiently small}, \tag{7}$$

$$P_m \text{ is real}, \tag{8}$$

$$\xi \text{ real} \neq 0 \Rightarrow \sum_1^n \left| \frac{\partial P_m(x,\xi)}{\partial \xi_k} \right| > 0. \tag{9}$$

The solvability of (6) was a great breakthrough in the theory and the conditions (7) and (9) but perhaps not (8) of reasonable generality. There were, however, surprises to come. Let us weaken (6) to

$$Pu = v \text{ has a solution (possibly a distribution) for every } v \in C_0. \tag{10}$$

By an ingenious but very special reasoning Hans Lewy proved in 1957 that the operator

$$D_1 + iD_2 + i(x_1 + ix_2) D_3$$

does not have this property. An almost complete investigation of the situation was given by Hörmander in 1960. Rephrasing (10) in the form of an inequality he got the following result. Put

$$C(x, D) = (P(x, D)\overline{P}(x, D) - \overline{P}(x, D)P(x, D))_{2m-1}.$$

This operator only depends on P_m and vanishes if P_m is real. Then one has

$$(10) \Rightarrow (P_m(x,\xi) = 0, \quad \xi \text{ real} \Rightarrow C(x,\xi) = 0). \tag{11}$$

Furthermore, this statement almost has a converse. We say that P is principally normal if there exists a polynomial $Q_{m-1}(x, \xi)$ of degree $m-1$ in ξ such that

$$C(x, \xi) = P_m(x, \xi)\overline{Q}_{m-1}(x, \xi) + \overline{P}_m(x, \xi)Q_{m-1}(x, \xi).$$

This property implies the right side of (11) and Hörmander proved that

$$P \quad \text{principally normal} \tag{12}$$

together with (7) and (9) implies the solvability of (6). The discovery of these results and their proofs is a first-class achievement. Recently he has given a global version of the theory.

I will finish by touching upon the problem of unique continuation which is the following. Consider a regular surface $S: s(x) = 0$ in R^n and a solution u of $Pu = 0$ which vanishes on one side of S. Does it vanish on the other side? If the coefficients of P are analytic the answer is yes provided S is non-characteristic for P, i.e.

$$P_m(x, \operatorname{grad} s(x)) \neq 0. \tag{13}$$

This is an old result by Holmgren. The same problem for non-analytic coefficients is much harder and there are counterexamples to show that (13) is not enough. The first positive results are due to Carleman for two variables and to Calderón in the general case. Hörmander has clarified the situation considerably by inventing a sufficient convexity condition bearing on S and P which also comes close to being necessary.

I hope that this sketch has given you an idea of the power and the drive that characterizes Hörmander's work.

REFERENCE

HÖRMANDER, L., *Linear Partial Differential Operators*, Springer 1963.

THE WORK OF JOHN W. MILNOR

By HASSLER WHITNEY

To aid in understanding the significance of Milnor's work, I will first say a few words about the recent history of algebraic topology. In the early thirties, the subject seemed to have reached a certain level of completeness in its basic methods and results; the famous textbooks of Lefschetz, of Seifert and Threlfall, and of Alexandroff and Hopf, gave a very good picture of the field. The term "algebraic" had not yet been applied. Then in 1935, the sudden explosion of "cohomology theory", with its various kinds of applications, provided a great impetus to research. In developing these applications to new heights, various algebraic problems arose; one could often write out a formula which described a situation, but could not understand the geometric meaning of the formula. In the forties, powerful and general algebraic machinery came into being; this enabled problems, which had before seemed hopelessly complex, to be answered in relatively simple algebraic terms. The subject of "algebraic topology" was in full swing.

The new algebraic methods now began to grow by themselves; new concepts, such as exact sequences, sheaves, homological algebra, spread out not only through topology but through neighboring domains of algebra and analysis, where they have had extraordinary success. As pointed out by H. Hopf, in presenting the work of the Fields Medalist R. Thom at the last Congress at Edinburgh, the geometric point of view tended to be swamped in the algebraic machinery. Papers in algebraic topology had commonly the appearance of being pure algebra. Then in the mid fifties, some great discoveries of a more geometric nature took place, which brought on a resurgence of the geometric point of view. The definition of cobordism and its basic properties and applications by Thom in 1954 was one such discovery; for this he was awarded a Fields Medal in 1958, as mentioned above. In 1956, the mathematical world was astounded by Milnor's proof that the 7-dimensional sphere S^7 was capable of several differential structures. This at first might have seemed like an isolated fact; but through Milnor and others it has led to a vast field of work, which has now acquired a name: *differential topology*.

To explain this work, I must give some definitions. A *differential manifold* is a topological manifold with given local coordinate systems, such that the transformation from one coordinate system to an overlapping one is differentiable. I (and some others) have purposely altered the standard term, "differentiable manifold", precisely because Milnor's result shows that one must tell, not merely whether a manifold is "differentiable", but which differential structure one has in mind. (Recently M. Kervaire has shown that there are topological manifolds which admit no differential structure at all.) A homeomorphism between two topological manifolds is a one–one correspondence between them which is continuous in both directions; a *diffeomorphism* is a homeomorphism that is differentiable in both directions.

We may now state Milnor's result: There are differential manifolds which are homeomorphic to the 7-sphere S^7, but not diffeomorphic to it.

A first reaction to such a theorem might be: What is the use in showing that S^7 can be distorted to such a degree that the original differential structure is lost, but a new one is taken on, distinct from the first? The fact is, the new manifolds are constructed not by pathological means, but by simple and well known methods; the actual differential manifolds had certainly been considered, but one had not suspected that they were homeomorphic to the sphere, and yet had different differential structures.

The manifolds are defined as follows. Start with the Cartesian product $S^4 \times S^3$. Cut S^4 apart at its equator S^3, thus cutting the product manifold apart. Now glue the two resulting bounded manifolds together again at their boundaries, identifying (u,v) with (u,v'), where $v' = u^h v u^j$; the multiplication here is quaternion multiplication in $S^3 \subset R^4$. Taking $h+j=1$, $h-j=k$, for any odd integer k, gives a manifold M_k^7; it is the total space of a sphere bundle over S^4, and has an obvious differential structure. It is easy to show that there is a differentiable real valued function on M_k^7 which admits just two critical points, these being nondegenerate; from this it follows that M_k^7 is homeomorphic to S^7.

An invariant λ for closed oriented differential manifolds M^7 is now defined, as follows: As part of Thom's cobordism theory, M^7 is the boundary of a differential manifold B^8. (For the M_k^7, one may take B_k^8 to be the total space of the associated disk bundle.) Now $\lambda(M^7)$ is defined as a certain function of the "index" and the first Pontrjagin class of B^8; it is an integer mod 7, independent of the choice of B^8. Moreover, $\lambda(M_k^7) \equiv k^2 - 1 \bmod 7$, and $\lambda(S^7) \equiv 0$; hence M_3^7 for instance is homeomorphic but not diffeomorphic to S^7, and it thus may be considered as defining a new differential structure on the topological manifold S^7.

Milnor later extended this work to a study of spheres of dimension $4m-1$, with $m>2$. The 31-sphere, for example, was shown to admit more than sixteen million distinct differential structures.

The next job in building up a theory of these structures was to form them into some sort of mathematical system. Milnor showed that an abelian group θ^n could be constructed from them, for each dimension n. We give a brief description of θ^n. The *connected sum* $M_1 \# M_2$ of two differential n-manifolds M_1 and M_2 is formed by removing a small n-cell from each, and then gluing together the resulting boundaries, with proper regard to orientation. Two closed differential manifolds M and M' are *h-cobordant* (homotopy cobordant) if the disjoint sum $M+(-M')$ is the boundary of some differential manifold W, such that both M and M' are deformation retracts of W; this is an equivalence relation. Now the group θ^n is the abelian group of h-cobordism classes of differential manifolds which are homotopy n-spheres (that is, have the same homotopy type as S^n), under the connected sum operation.

Through some deep work of S. Smale and others, it is known that the elements of θ^n correspond precisely to the distinct diffeomorphism classes of S^n. Hence, for instance, the order of θ^n is the number of distinct differential structures on S^n. These groups have wide importance; for instance, they serve as coefficient groups in the study of obstructions to the construction of differential structures on a manifold. They play a role, relative to differentiability, similar to the role of the homotopy groups in more purely topological questions.

Kervaire and Milnor, in a joint study, have shown that the groups θ^n

are always finite, except that the structure of θ^3 is unknown. (If the Poincaré conjecture were proved, it would follow that θ^3 is trivial.) They have found the orders of many of these groups; for instance:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
order of θ^n	1	1	?	1	1	1	28	2	8	6	992	1	3	2	16256

The order of θ^{4m-1} is given in terms of the order of the stable homotopy group $\pi_{k+4m-1}(S^k)$ (k large) and the Bernoulli number B_m.

We must mention the recent solution by Milnor of two outstanding problems of topology; in each case the solution was made possible by deep work of someone else. First, when R. Bott proved a theorem about the divisibility of the highest Pontrjagin class of the $4m$-sphere, Milnor saw that, combining this with results of Wu Wen-Tsün, it followed that the n-sphere is parallelizable only if $n=1$, 3 or 7. (Kervaire and Hirzebruch also discovered this independently.) A further consequence of this is that there exists a division algebra of rank n over the real numbers if and only if $n=1$, 2, 4 or 8. (In these cases, examples are: the real numbers, the complex numbers, the quarternions, and the Cayley numbers.)

Second, the so-called "Hauptvermutung" for complexes, that homeomorphic polyhedra have isomorphic cell subdivisions, was formulated in a paper by H. Tietze in 1908 (it was conjectured earlier). In this paper, he studied the "lens spaces" $L(p,q)$. These are manifolds, defined as follows. Set $\omega=e^{2\pi i/p}$. Taking the unit sphere S^3 in the space of pairs (z_1,z_2) of complex numbers and identifying any two points (z_1,z_2), $(\omega z_1,\omega^q z_2)$ of S^3 defines $L(p,q)$. Tietze showed that p is a topological invariant. With the help of a very recent theorem of B. Mazur on the existence of a diffeomorphism between certain product spaces, Milnor shows that $L(p,q)\times R^n$ is diffeomorphic to $L(p,q')\times R^n$ provided that $\pm qq'$ is a quadratic residue mod p, and $n>3$. Now construct the finite complex X_q from the product $L(7,q)\times\sigma^n$, σ^n being an n-simplex, by adjoining a cone over the boundary $L(7,q)\times\partial\sigma^n$. Then for $n\geqslant 3$, X_1 is homeomorphic with X_2, as can be proved from the result stated above. Now the Hauptvermutung is *disproved* by showing that X_1 and X_2 have no isomorphic cell subdvisions. This is accomplished by slightly generalizing the torsion invariant of Reidemeister, Franz and de Rham, and following some methods developed by J. H. C. Whitehead.

One possible peculiarity of differential structures appears as a corollary of the work just mentioned: The bounded manifolds $L(7,1)\times D^5$ and $L(7,2)\times D^5$ (where D^5 is the unit 5-disk) are not diffeomorphic, but their interiors are.

To look ahead, Milnor is now developing a theory of "microbundles", which injects a new and basic tool into the subjects discussed above. You will hear about this from him in his address at this Congress.

In conclusion, it is apparent that differential topology is a strong young field, with a bright present and a great outlook, and that its vitality is largely due to the fine achievements of Milnor. It gives me especial pleasure to report on this; in the thirties I was much concerned with the relations between differential and topological properties, and I felt that my own contributions were merely some beginnings of what might come at some time in the future. The future has arrived!

ONE-HOUR ADDRESSES

TEICHMÜLLER SPACES

By LARS V. AHLFORS(1)

1. Historical remarks

Riemann's classical problem of moduli is not a problem with a single aim, but rather a program to obtain maximum information about a whole complex of questions which can be viewed from several different angles. At this congress other reports will be presented which deal with the remarkable progress that has been achieved by purely algebraic methods, as well as by transcendental methods with an algebraic background. The present lecture reports on the results obtained by an analytic method which is based on the geometric idea of quasiconformal mapping.

This analytic treatment has its roots in the work of O. Teichmüller. He discovered that a certain extremal problem of quasiconformal mapping is intimately connected with the quadratic differentials on a Riemann surface. On the other hand, the number of linearly independent quadratic differentials is the same as the number of moduli. Teichmüller surmised, but did not establish, a link between moduli and quasiconformal mappings. The results on moduli that he achieved later were based on a different method which did not have the simplicity of his original approach.

Teichmüller's work was too sketchy to serve as a satisfactory theory. During the 1950's his results were confirmed and reorganized by H. Rauch, L. Bers, and the speaker. Through these efforts it was firmly established that the space of moduli has a natural complex analytic structure. As the work progressed, strong connections became apparent with Beltrami's classical theory of conformal mapping of surfaces with a Riemannian metric. Much more delicate existence theorems were needed, but it was found that they had already been proved by C. Morrey for different purposes. Ideas of I. N. Vekua and Boyarskii led to important simplifications of these existence proofs.

2. Teichmüller's covering space

To study the moduli is to study the space whose points are the classes of conformally equivalent Riemann surfaces. In its natural topology this space is not a manifold, and for this reason Teichmüller preferred to consider a covering space obtained by a weaker equivalence relation.

The formal definition of Teichmüller's space is as follows: Let W_0 be a compact Riemann surface. Consider pairs (W,f) where W is another Riemann surface and f is a sense-preserving homeomorphic mapping of W_0 on W. Write $(W_1,f_1)\sim(W_2,f_2)$ if there exists a conformal homeomorphism h: $W_1\to W_2$ such that $f_2^{-1}hf_1$ is homotopic to the identity. The set of pairs (W,f), identified under this relation, is the Teichmüller space $T(W_0)$. Given a pair (W_1,f_1) there is a natural identification of $T(W_1)$ and $T(W_0)$, namely by the correspondence $(W,f)\leftrightarrow(W,ff_1)$. We regard the Teichmüller spaces as the same, but referred to different frames. In this sense all surfaces of fixed genus g yield the same space T_g.

(1) This research was supported in part by the Air Force Office of Scientific Research.

If $W_1 = W_0$ we obtain a mapping $\beta:(W,f)\to(W,ff_1)$ of $T(W_0)$ on itself. Its fixpoints correspond to surfaces with nontrivial conformal selfmappings. The mappings β form the "modular" group M, and $T(W_0)/M$ may be regarded as the space of moduli in the original sense.

For many purposes it is preferable to introduce another interpretation. A surface W_0 of finite genus $g>1$ can be represented as a quotient space H/Γ_0 where H is the upper halfplane and Γ_0 is a Fuchsian group. To a pair (W,f) corresponds another Fuchsian group Γ and an isomorphism ϑ from Γ_0 to Γ. The relation $(W_1,f_1)\sim(W_2,f_2)$ means that ϑ_1 and ϑ_2 are conjugate isomorphisms. In other words, there exists a linear transformation B such that $\vartheta_2(A)B = B\vartheta_1(A)$ for all $A\in\Gamma_0$. With this equivalence relation $T(W_0)$, or more simply $T(\Gamma_0)$, is a set of equivalence classes of isomorphisms ϑ. A change of frame is represented by a correspondence $\vartheta\leftrightarrow\vartheta\vartheta_1$. The modular mappings correspond to automorphisms ϑ_1 of Γ_0.

This construction can be generalized to arbitrary discontinuous groups Γ_0. An important case is that of quasifuchsian groups which are discontinuous on two complementary Jordan regions.

3. Global coordinates

The most effective technique to study $T(W_0)$ is by deformation of the conformal structure. A Beltrami differential on W_0 is a differential form $\mu d\bar{z}/dz$ of weight $(-1,1)$. If μ is sufficiently regular, and $|\mu|<1$, it defines a new conformal structure on W_0, namely by the local fundamental forms $ds^2 = |dz+\mu d\bar{z}|^2$. If the surface W_0 with this structure is denoted by W_0^μ, the pair (W_0^μ, e), where e is the identity mapping, represents an element of $T(W_0)$. Provided that the regularity conditions are not too strict, it is easy to show that every point of $T(W_0)$ can be represented in this form. The proper degree of generality seems to be to require that μ is merely measurable, with norm $\|\mu\|_\infty<1$. If $B(W_0)$ is the linear space of all measurable Beltrami differentials, we are thus in a position to identify $T(W_0)$ with the unit ball $B_1(W_0)$ modulo a certain equivalence relation $\mu_1\sim\mu_2$.

This identification can be made quite explicit. Let W_0 be represented as H/Γ_0 where H is again the upper halfplane and Γ_0 is a Fuchsian group of the first kind. The Beltrami differentials become functions $\mu(z)$ on H which satisfy the relations

$$\mu(Az)\bar{A}'(z) = \mu(z)A'(z) \quad (A\in\Gamma_0). \tag{1}$$

Let μ be extended to the full plane by setting $\mu=0$ in the lower halfplane. A fundamental existence theorem, due to Morrey, and sometimes referred to as the generalized Riemann mapping theorem, asserts the existence and essential uniqueness of a homemorphism f^μ of the plane which satisfies the Beltrami equation

$$f^\mu{}_{\bar{z}} = \mu f^\mu{}_z \tag{2}$$

Here the derivatives are weak L^2-derivatives, and uniqueness is achieved by a suitable normalization, for instance by requiring that $0,1,\infty$ remain fixed.

It follows from (1) and (2) that

$$f^{\mu}(Az)=A^{\mu}f^{\mu}(z), \tag{3}$$

where A^{μ} is a linear transformation. The A^{μ} form a quasifushsian group Γ^{μ}, and the mapping: $A\to A^{\mu}$ is an isomorphism.

One sees very easily that $\mu_1\sim\mu_2$ if and only if $f^{\mu_1}=f^{\mu_2}$ on the real axes. Since they are analytic in the lower halfplane the equality remains valid there. Moreover, it follows from (3) that the Schwarzian derivative $\varphi^{\mu}=\{f^{\mu},z\}$ satisfies

$$\varphi^{\mu}(Az)A'(z)^2=\varphi^{\mu}(z), \tag{4}$$

which means that φ^{μ} is a quadratic differential with respect to Γ_0. Because of the normalization $\varphi^{\mu_1}=\varphi^{\mu_2}$ implies $f^{\mu_1}=f^{\mu_2}$ in $\overline{H}$, and hence $\mu_1\sim\mu_2$. As a result we are able to identify $T(W_0)$ with a certain subset $\Omega(\Gamma_0)$ of the space $Q(\Gamma_0)$ of quadratic differentials, $\Omega(\Gamma_0)$ being the image of $B_1(\Gamma_0)$ under the mapping $\mu\to\varphi^{\mu}$. In the case of T_g the complex dimension of $Q(\Gamma_0)$ is $3g-3$.

It is proved by fairly elementary considerations that $\Omega(\Gamma_0)$ is a bounded set which contains a neighborhood of the origin. In order to show that $\Omega(\Gamma_0)$ is open, and hence that $T(W_0)$ is a manifold, it is necessary to study the effect of a change of frame.

For this purpose, let $W_0=H/\Gamma_0$ be mapped on $W_1=H/\Gamma_1$ by a mapping F with complex dilatation ϱ, i.e. $F_{\bar z}=\varrho F_z$. We denote the dilatation of $f^{\mu}\circ F^{-1}$ by $\mu|\varrho$. It belongs to $B(\Gamma_1)$, and the correspondence $\mu\to\mu|\varrho$ induces a mapping $\varphi^{\mu}\to\varphi^{\mu|\varrho}$ of $\Omega(\Gamma_0)$ on $\Omega(\Gamma_1)$, for it is clear that $\mu_1\sim\mu_2$ implies $\mu_1|\varrho\sim\mu_2|\varrho$. It is practically evident that this mapping is bicontinuous, and it follows by the invariance of region that $\varphi^{\varrho}\in\Omega(\Gamma_0)$, which corresponds to $0\in\Omega(\Gamma_1)$, is an interior point. Hence $\Omega(\Gamma_0)$ is open.

But much more is true. The explicit formula

$$\mu|\varrho=\left(\frac{\mu-\varrho}{1-\bar\varrho\mu}\cdot\frac{F_z}{\overline{F_z}}\right)\circ F^{-1} \tag{5}$$

shows that $\mu|\varrho$ depends holomorphically on μ. It is also known that f^{μ}, and therefore φ^{μ}, is a holomorphic function of μ. From these facts it follows that $\varphi^{\mu}\to\varphi^{\mu|\varrho}$ is a holomorphic mapping. We conclude that T_g carries a complex analytic structure of dimension $3g-3$ which is independent of the frame.

This reasoning is essentially due to Bers. It is the simplest and shortest way to introduce a complex structure. Of course, it must be shown that this is the natural structure, for instance in the sense that the periods of the normal integrals become holomorphic functions on T_g. This is not a difficult task. There are two open problems: (1) to study the boundary of $\Omega(\Gamma_0)$, (2) to prove that $\Omega(\Gamma_0)$ is open even when $Q(\Gamma)$ is infinite dimensional.

There is an interesting application to the modular transformations. The correspondence $\varphi^{\mu}\to\varphi^{\mu|\varrho}$ is a cover transformation if $\Gamma_1=\Gamma_0$, in which case F maps H on itself. Suppose further that the origin is a fixpoint. Then F has the same boundary values as a linear transformation B which belongs to the

normalizer of Γ_0. It follows that $f^\mu = f^{\mu|\varrho} \circ B$ on $\bar{H}$. On taking Schwarzian derivatives we obtain

$$\varphi^\mu = (\varphi^{\mu|\varrho} \circ B)\, B'^2 \tag{6}$$

which is a linear transformation on $Q(\Gamma_0)$.

The subgroup of M which leaves the basepoint fixed can thus be represented as a group of linear transformations. When studying the singularities of $T(W_0)/M$ we are therefore merely required to investigate the quotient space of a finite dimensional linear space with respect to a finite linear group.

4. The metric on T_g

We regard T_g as a complex manifold and the representatives $T(W)$ as coordinate mappings. To study the differential geometry on T_g we have to determine the linear part of φ^μ.

To this end we introduce the Fréchet differentials $\dot{f}^\mu[\nu]$ and $\dot{\varphi}^\mu[\nu]$ of which the former is defined as

$$\dot{f}^\mu[\nu] = \lim_{t\to 0} \frac{1}{t}(f^{\mu+t\nu} - f^\mu). \tag{7}$$

For $\mu = 0$ the notation is simplified to $\dot{f}[\nu]$, $\dot{\varphi}[\nu]$.

It is immediate that $\dot{f}[\nu]_{\bar{z}} = \nu$ on H and O on $\bar{H}$. On taking the normalization into account it follows that

$$\dot{f}[\nu](z) = -\frac{1}{\pi}\int_H \left(\frac{1}{\zeta - z} - \frac{1-z}{\zeta} - \frac{z}{\zeta - 1}\right)\nu(\zeta)\, d\xi d\eta \quad (\zeta = \xi + i\eta). \tag{8}$$

Furthermore, the infinitesimal form of the relation $\{f^\mu, z\} = \varphi^\mu$ is

$$\dot{f}[\nu]''' = \dot{\varphi}[\nu], \tag{9}$$

valid on $\bar{H}$. On H we have therefore, by use of the symmetry,

$$\dot{\varphi}[\nu](z) = -\frac{6}{\pi}\int_H \frac{\bar{\nu}(\zeta)}{(\bar{\zeta} - z)^4}\, d\xi d\eta. \tag{10}$$

It is convenient to introduce the kernel

$$K(z, \bar{\zeta}) = (z - \bar{\zeta})^{-2} \tag{11}$$

and write (10) as

$$\dot{\varphi}[\nu](z) = -\frac{6}{\pi}\int_H K(z, \bar{\zeta})^2\, \bar{\nu}(\zeta)\, d\sigma(\zeta) \tag{12}$$

where $d\sigma$ is the area element.

One verifies that $\dot{\varphi}[\nu]$ is itself a quadratic differential. Let $N(\Gamma_0)$ be the subspace of $B(\Gamma_0)$ for which $\dot{\varphi}[\nu] = 0$. Then (12) represents an an-

tilinear one to one map of $B(\Gamma_0)/N(\Gamma_0)$ onto $Q(\Gamma_0)$. Indeed, there is a Bergman type identity

$$\varphi(z) = \frac{12}{\pi} \int_H K(z, \bar{\zeta})^2 \varphi(\zeta) (\operatorname{Im} \zeta)^2 \, d\sigma(\zeta) \tag{13}$$

which shows that $\nu(z) = -2\bar{\varphi}(z) y^2$ corresponds to $\dot{\varphi}[\nu] = \varphi$.

Choose a basis $\nu_1, \ldots, \nu_n$ $(n = 3g - 3)$ of $B(\Gamma_0)$ modulo $N(\Gamma_0)$. The mapping

$$\tau = (\tau_1, \ldots, \tau_n) \to \varphi^{\bar{\tau}_1 \nu_1 + \ldots + \bar{\tau}_n \nu_n} \tag{14}$$

turns out to be holomorphic, and τ can be used as local coordinate. A function on $T(W_0)$ can be written as $G(\bar{\tau}_1 \nu_1 + \ldots + \bar{\tau}_n \nu_n)$ and it is holomorphic if $\dot{G}(\mu)[\nu]$ is antilinear in ν. The derivatives are

$$\frac{\partial G}{\partial \tau_k} = \dot{G}(\mu)[\nu_k] \tag{15}$$

at $\mu = \bar{\tau}_1 \nu_1 + \ldots + \bar{\tau}_n \nu_n$. There is thus an antilinear correspondence between the tangent plane and $B(\Gamma_0)/N(\Gamma_0)$, or a linear correspondence between tangents and quadratic differentials.

We wish to introduce a Hermitian metric

$$ds^2 = \sum g_{\alpha\bar{\beta}}(\tau) \, d\tau_\alpha \, d\bar{\tau}_\beta. \tag{16}$$

To do this we set, first of all,

$$g_{\alpha\bar{\beta}}(0) = \langle \nu_\beta, \nu_\alpha \rangle = -\int_{H/\Gamma_0} \nu_\beta \dot{\varphi}[\nu_\alpha] \, d\sigma, \tag{17}$$

or more explicitly, by (12),

$$g_{\alpha\bar{\beta}}(0) = \frac{6}{\pi} \int_{H/\Gamma_0} \int_H K(z, \bar{\zeta})^2 \nu_\beta(z) \, \bar{\nu}_\alpha(\zeta) \, d\sigma(z) \, d\sigma(\zeta). \tag{18}$$

It is to be noticed that $\langle \nu_\beta, \nu_\alpha \rangle$ depends only on the equivalence classes of ν_α and ν_β, and is thus an inner product on $B(\Gamma_0)/N(\Gamma_0)$. In (18), the domain of integration $H/\Gamma_0 \times H$ may be replaced by $H \times H/\Gamma_0$, which proves the Hermitian symmetry. Finally, if we use the fact that $\nu_\beta \equiv -2\dot{\bar{\varphi}}[\nu_\beta]$ mod $N(\Gamma_0)$ we obtain

$$g_{\alpha\bar{\beta}}(0) = 2 \int_{H/\Gamma_0} \dot{\varphi}[\nu_\alpha] \, \dot{\bar{\varphi}}[\nu_\beta] \, y^2 \, d\sigma(z). \tag{19}$$

The positive definite character is now evident. Incidentally, (19) is Petersson's inner product of quadratic differentials. Its use in this connection was originally proposed by A. Weil.

To determine $g_{\alpha\bar{\beta}}(\tau)$ all we need is to require invariance under the mapping $\mu \to \mu \,|\, \varrho$ with $\varrho = \bar{\tau}_1 \nu_1 + \ldots + \bar{\tau}_n \nu_n$. For any function G on $B(\Gamma_0)$ let us define G_ϱ on $B(\Gamma^\varrho)$ by

$$G(\mu) = G_\varrho(\mu \,|\, \varrho). \tag{20}$$

It follows from (5) that

$$\dot{G}(\varrho)[\nu] = \dot{G}_\varrho[L_\varrho \nu], \tag{21}$$

where

$$L_\varrho \nu = \left(\frac{\nu}{1-|\varrho|^2}\frac{F^\varrho_z}{\overline{F}^\varrho_{\bar{z}}}\right) \circ (F^\varrho)^{-1}. \tag{22}$$

This correspondence between the tangent spaces shows that we must have

$$g_{\alpha\bar{\beta}}(\tau) = \langle L_\varrho \nu_\beta, L_\varrho \nu_\alpha \rangle, \tag{23}$$

where the inner product is now with respect to H/Γ^ϱ.

By use of (18) we have thus

$$g_{\alpha\bar{\beta}}(\tau) = \frac{6}{\pi}\int_{H/\Gamma^\varrho}\int_H K(z,\bar{\zeta})^2\, L_\varrho \nu_\beta(z)\, \overline{L_\varrho \nu_\alpha}(\zeta)\, d\sigma(z)\, d\sigma(\zeta) \tag{24}$$

and a change of integration variables brings this to the form

$$g_{\alpha\bar{\beta}}(\tau) = \frac{6}{\pi}\int_{H/\Gamma_0}\int_H K_\varrho(z,\bar{\zeta})^2\, \nu_\beta(z)\, \bar{\nu}_\alpha(\zeta)\, d\sigma(z)\, d\sigma(\zeta), \tag{25}$$

where we have set

$$K_\varrho(z,\bar{\zeta}) = K(F^\varrho(z), \overline{F}^\varrho(\zeta))\, F^\varrho_z \overline{F}^\varrho_{\bar{\zeta}}. \tag{26}$$

5. Curvature properties

The comparatively simple structure of formula (25) makes it possible to find explicit expressions for the successive derivatives of the coefficients $g_{\alpha\bar{\beta}}$ and thereby to study the curvature properties of the metric. Naturally, the computations are rather involved, and in the frame of this lecture it is possible to give only an outline of the results. For details I refer to my paper in the *Journal d'Analyse*.(1)

The first step is to find the derivatives of K_ϱ. If we note that

$$\dot{F}[\nu](z) = \dot{f}[\nu](z) + \dot{\bar{f}}[\nu](\bar{z}), \tag{27}$$

it follows easily that

$$\frac{\partial F}{\partial \tau_\gamma} = \dot{f}[\nu_\gamma]. \tag{28}$$

After some computation one obtains

$$\frac{\partial}{\partial \tau_\gamma} K_\varrho(\zeta_1, \bar{\zeta}_2) = -\frac{1}{\pi}\int_H K_\varrho(\bar{z}, \bar{\zeta}_2)\, K_\varrho(\bar{z}, \zeta_1)\, \bar{\nu}_\gamma(z)\, d\sigma(z). \tag{29}$$

Here $K_\varrho(z,\zeta)$ is formed exactly as $K_\varrho(z,\bar{\zeta})$, but without conjugations, and the integral has to be interpreted as a principal value.

(1) Vol. 9, 1961.

With the help of (25) and (29) we now derive, at $\tau=0$,

$$\frac{\partial g_{\alpha\bar{\beta}}}{\partial \tau_\gamma} = -\frac{12}{\pi^2}\int_{H/\Gamma_0}\int_H\int_H K(\zeta_1,\bar{\zeta}_2)K(\bar{\zeta}_2,\bar{\zeta}_3)K(\zeta_1,\bar{\zeta}_3)\nu_\beta(\zeta_1)\bar{\nu}_\alpha(\zeta_2)\bar{\nu}_\gamma(\zeta_3)d\sigma_1 d\sigma_2 d\sigma_3. \tag{30}$$

We point out that in this formula one of the integration variables ranges over a fundamental region, and the others over the whole halfplane. It never matters which one we let range over the fundamental region.

The integral in (30) is symmetric in α and γ. This means that the metric is Kählerian.

Higher derivatives can be obtained in the same way. Indeed, (30) is valid also for $\tau \neq 0$, provided that K is replaced by K_ϱ. By use of (29) and a similar formula for the derivatives of $K_\varrho(\zeta_1,\zeta_2)$ it is evidently possible to express $\partial^2 g_{\alpha\bar{\beta}}/\partial\tau_\gamma\partial\bar{\tau}_\delta$ as an integral over $H/\Gamma_0 \times H^3$ and the integrand will be a sum of three terms, each containing a product of four K's.

A paper of S. Bochner contains complete formulas for computing curvatures of a Kählerian metric. One obtains expressions for the Ricci curvatures and curvatures of holomorphic sections which show that they are strictly negative. Presumably, it could be shown that the scalar curvature is not constant, but nobody has carried out the computations.

ARITHMETIC PROPERTIES OF LINEAR ALGEBRAIC GROUPS

By ARMAND BOREL

The theory of Lie groups in the large has been developed chiefly from two points of view: the transcendental one, shaped by the work of H. Weyl and E. Cartan, which emphasizes maximal compact subgroups, Riemannian symmetric spaces, and is a meeting ground for Lie algebra methods, differential geometry, topology; and the more recent point of view of algebraic groups, based on the use of algebraic geometry, which has produced results in arbitrary characteristic but has also given new information and fruitful ideas in the classical case. All these, combined, have provided a suitable framework for a further development, linking arithmetic and algebraic groups, which is the subject matter of the present report.

The impetus for the recent work in that area came to a large extent from two sources. First, it was noticed some years ago that the formalism of adeles and ideles introduced in algebraic number theory by Chevalley could also be applied to linear algebraic groups. This allowed one to give new and simple formulations to a number of classical results, in particular on quadratic forms, which made sense for more general algebraic groups. They suggested therefore new problems as well as the possibility of a more unified treatment of the cases considered so far. The other incentive has been the wish to develop in a more general setting the theory of automorphic functions of several variables.

Important in both cases is the study of fundamental domains for certain discrete subgroups of real algebraic groups, called arithmetic groups or groups of units, to which I shall devote about the first half of this talk. The second part will be concerned with problems involving also p-adic algebraic groups or adele groups. Automorphic functions are related to these questions, but my time is limited and, more important, some of the main developments in that area will be discussed in other lectures, notably in those of Selberg [24] and Gel'fand [11]. I shall therefore leave them aside and concentrate on more algebraic questions, or on the more algebraic aspects of these questions.

1. Groups of units

I would like now first to define the groups of units in general, and then give a few examples and state some of their properties.

A subgroup G of $\mathrm{GL}(n, \mathbf{C})$, the group of $n \times n$ complex invertible matrices, is *algebraic* if it is the set of all invertible matrices whose coefficients annihilate some collection of polynomials in n^2 indeterminates, with complex coefficients; it is said to be defined over a subfield k of $\mathbf{C}$ if those polynomials may be chosen so as to have their coefficients in k. The intersection of the fields of definition is also a field of definition. If B is a subring of $\mathbf{C}$, then G_B denotes the group of elements of G which have their coefficients in B

and whose determinant is a unit of B, i.e. is invertible in B. In the sequel, unless otherwise said, G will stand for an algebraic subgroup of $\mathbf{GL}(n, \mathbf{C})$ which is defined over the field $\mathbf{Q}$ of rational numbers. Its *group of units* is then $G_{\mathbf{Z}}$, the group of integral matrices with determinant ± 1 contained in G. A group of units is thus attached to a matric group, its definition presupposes the choice of a basis. If we perform a rational change of coordinates or, more generally, if we replace G by its image under a faithful rational representation defined over $\mathbf{Q}$, then the new group of units is in general different from the first one, but is at any rate commensurable with it. Thus the commensurability class of groups of units has a more intrinsic meaning. I recall that two subgroups of a group are commensurable if their intersection has finite index in both.

The group $G_{\mathbf{R}}$ of real matrices of G is a real Lie group (i.e. a real analytic manifold endowed with a group structure for which the product and the inverse are real analytic functions of their arguments) with finitely many connected components. Therefore, its maximal compact subgroups are conjugate to each other by inner automorphisms and the quotient $D = K \backslash G_{\mathbf{R}} = \{Kg,\ g \in G_{\mathbf{R}}\}$ of $G_{\mathbf{R}}$ by one of them is homeomorphic to a euclidean space. If G is semi-simple, i.e. if it has no infinite invariant commutative subgroup or, equivalently, if its Lie algebra is the direct product of its simple ideals, then D is a Riemannian symmetric space with negative curvature without flat component; and, conversely, every such space can be obtained in this way.

The group $G_{\mathbf{Z}}$ acts by right translations on $G_{\mathbf{R}}$, D, and is *properly discontinuous*: this means that every compact set meets only a finite number of its translates by $G_{\mathbf{Z}}$; in particular the set of transforms of a given point has no accumulation point. Our problem is then to find suitable sets of representatives for the orbits of $G_{\mathbf{Z}}$ in $G_{\mathbf{R}}$ or in D. These are essentially equivalent questions; for the proofs, it is often more convenient to work in $G_{\mathbf{R}}$, but the consideration of D brings us closer to the classical reduction theory. I shall therefore give the preference to D in this lecture.

2. Examples

(*a*) $G = \mathbf{GL}(n, \mathbf{C})$, $K = \mathbf{O}(n)$, the group of real $n \times n$ orthogonal matrices, and D is the space of positive non-degenerate real symmetric $n \times n$ matrices, on which $G_{\mathbf{R}} = \mathbf{GL}(n, \mathbf{R})$ acts in the standard fashion by:

$$A \to {}^tX \cdot A \cdot X \quad (A \in D,\ X \in \mathbf{GL}(n, \mathbf{R}),\ {}^tX \text{ transpose matrix of } X).$$

Here $G_{\mathbf{Z}}$ is the group of integral matrices with determinant ± 1, and its orbits are just the *classes of positive non-degenerate quadratic forms*. The construction of sets of representatives is the task of the reduction theory of postive quadratic forms which goes back to Gauss for $n = 2$, to Hermite [12] for $n \geqslant 3$.

(*b*) Let F be a rational non-degenerate quadratic form on $\mathbf{C}^n$ and $G = 0(F)$ the orthogonal group of F, i.e. the group of linear transformations leaving F invariant. $G_{\mathbf{Z}}$ is then classically called the group of units of F. In this case, D is the space of majorizing forms of F, in the sense of Hermite [12]: these are the minimal elements among the positive forms S which majorize

F in the obvious sense (i.e. verify $|F(x,x)| \leqslant S(x,x)$ for every $x \in \mathbf{R}^n$); they can also be defined as those positive quadratic forms whose matrices satisfy the relation $F \cdot S^{-1} \cdot F = S$. The maximal compact subgroups of $G_{\mathbf{R}}$ are the subgroups of $G_{\mathbf{R}}$ which leave invariant one of those majorizing forms.

(c) G is the symplectic group $\mathbf{Sp}(2n)$, the group of $2n \times 2n$ matrices leaving invariant the antisymmetric bilinear form

$$F(x,y) = \sum_{1 \leqslant i \leqslant n} (x_i \cdot y_{n+i} - y_i \cdot x_{n+i}).$$

The group $G_{\mathbf{Z}}$ is then Siegel's modular group and D is the generalized upper half-plane, the space of complex $n \times n$ symmetric matrices with positive non-degenerate imaginary part [28].

3. Fundamental sets for groups of units

A fundamental domain for a group acting on a set is, strictly speaking, a subset containing exactly one representative of each orbit. Here, we shall be concerned only with suitable approximations to fundamental domains in this strict sense, to be called fundamental sets.

A subset $\Omega \subset D$ is a *fundamental set* for $G_{\mathbf{Z}}$ if first it meets each orbit of $G_{\mathbf{Z}}$, i.e. if:

(i) $\Omega \cdot G_{\mathbf{Z}} = D$.

A second natural requirement is that Ω should intersect only a finite number of its translates $\Omega \cdot x$ by elements of $G_{\mathbf{Z}}$; however, when dealing with arithmetic groups, it is important to have at one's disposal the stronger condition:

(ii) For any $a \in G_{\mathbf{Q}}$, the set of $x \in G_{\mathbf{Z}}$ for which $\Omega \cdot a \cap \Omega \cdot x$ is not empty is finite,

and this will be our second condition for fundamental sets.

In our first example above, we may take as fundamental sets the so-called *Siegel domains*. The Siegel domain $\mathfrak{S}_{t,u}$ $(t,u>0)$ in the space of positive $n \times n$ real symmetric matrices is the set of products ${}^tN \cdot A \cdot N$, where $N = (n_{ij})$ is upper triangular, with ones in the diagonal, and $|n_{ij}| \leqslant u$ $(i<j)$, and where A is diagonal, with positive diagonal entries $a_1, \ldots, a_n$ verifying $a_i \leqslant t \cdot a_{i+1}$ $(1 \leqslant i \leqslant n-1)$. It satisfies our second condition by a well known result of Siegel [26]; that (i) is fulfilled when $t \geqslant 4/3$, $u \geqslant 1/2$ is implicit in Hermite's work [12], and made explicit in [15].

The existence of closed or open fundamental sets in the general case, which have a finite volume with respect to an invariant measure when G is semi-simple, (or more generally when the identity component of G has no non-trivial rational character defined over $\mathbf{Q}$) is proved in a paper written jointly by Harish-Chandra and myself [7]. The crucial case is when G is semi-simple, for which the construction of fundamental sets generalizes in a way the procedure used by Hermite [12] in the case of indefinite rational quadratic forms (our second example): D being realized as a submanifold of the space D' of all positive non-degenerate real symmetric matrices of some degree n', these fundamental sets are the intersection of D with a finite number of translates, by elements of $\mathbf{GL}(n',\mathbf{Z})$, of a Siegel domain of D'.

The paper [7] also gives a necessary and sufficient condition under which $D/G_{\mathbf{Z}}$ is compact. When G is semi-simple, the condition, which had been

conjectured by Godement, is that either $G_{\mathbf{Z}}$ or $G_{\mathbf{Q}}$ consists of semi-simple matrices (matrices which can be diagonalized over the complex numbers). In our second example above, this is equivalent to the non-existence of non-trivial rational zeros of F. From this it can be deduced that every Riemannian symmetric space with negative curvature has discontinuous groups of isometries with relatively compact fundamental domains which act freely (i.e. no element different from the identity has fixed points). In other words, every such space has compact Clifford-Klein forms [32]. Another proof of Godement's conjecture is given in [33].

4. Fundamental sets with cusps

The construction of fundamental sets in semi-simple groups defined over $\mathbf{Q}$ briefly mentioned previously depends chiefly on $G_{\mathbf{R}}$ and has comparatively little to do with $G_{\mathbf{Q}}$ or $G_{\mathbf{Z}}$. However, starting from this first approximation to fundamental domains, it is possible to construct a second, and usually better, one when $D/G_{\mathbf{Z}}$ is not compact, by taking also into account some properties of $G_{\mathbf{Q}}$. I have in mind here mainly the facts which center around the so-called Bruhat lemma, or rather around its generalization to rational points over non-algebraically closed fields, proved independently by Tits and myself ([5], [30]).

These new fundamental sets are better suited to the study of automorphic forms and of the compactification of $D/G_{\mathbf{Z}}$. They are the union of finitely many unbounded sets which behave more or less like the parabolic cusps of the fundamental domains for fuchsian groups in the upper half-plane, and which shall also be called cusps. These cusps are the translates by elements of $G_{\mathbf{Q}}$ of subsets which are defined roughly in the same way as the Siegel domains in $G_{\mathbf{R}}$ [7; § 4], but with the maximal subgroups which can be put in triangular or in diagonal form over $\mathbf{R}$ replaced by subgroups enjoying similar properties with respect to $\mathbf{Q}$ (see [5] for a precise description).(1)

The cusps correspond to the equivalence classes modulo $G_{\mathbf{Z}}$ of the maximal unipotent subgroups of $G_{\mathbf{Q}}$.(2) In some cases, they have other interpretations. For instance, if G is the orthogonal group of a rational quadratic form F, and if the integral points form a maximal lattice for F, they correspond to the equivalence classes modulo $G_{\mathbf{Z}}$ of maximal isotropic subspaces defined over $\mathbf{Q}$. If $G_{\mathbf{Z}}$ is the Hilbert-Blumenthal group associated to a number field k, they correspond to the ideal classes of k.

These fundamental sets can be used to show that the compactification of $D/G_{\mathbf{Z}}$ associated to a linear representation of G by Satake's procedure [23] consists of finitely many quotient spaces of the same type if the representation is defined over $\mathbf{Q}$. This result is valid for any semi-simple group defined over $\mathbf{Q}$ and yields only topological compactifications. It does not say whether some of those compactifications are normal complex analytic spaces when D is a bounded domain in the space of several complex variables. It does not seem too unlikely that this or related questions on automorphic functions

(1) It seems rather likely that this implies that the arithmetic groups are regular in the sense of Gel'fand [11].

(2) A matrix group is unipotent if it consists of matrices all of whose eigenvalues are equal to one; it can then be put in triangular form and is nilpotent.

can be handled with the help of the above fundamental sets and of certain results of Harish-Chandra (*Amer. J. Math.*, 78 (1956), 564–628), either by the Andreotti-Grauert method [2], or by showing the validity of certain conditions which are roughly analogous to the normality conditions discussed by Pyatetski-Šapiro in his book on classical domains and automorphic functions [22]. However, this has not yet been done, as far as I know.

5. Some finiteness theorems for groups of units

I would like now to state some properties of groups of units whose proofs make use of fundamental sets, although their statements do not.

(*a*) $G_{\mathbf{Z}}$ is finitely generated, finitely presented, and its finite subgroups form a finite number of conjugacy classes.

(Since only the finite generation of $G_{\mathbf{Z}}$ is proved in [7], I sketch the proof of the two other properties: the subgroup N of $G_{\mathbf{Z}}$ which acts trivially on D is finite, being contained in K; it is therefore sufficient to prove them for $G_{\mathbf{Z}}/N$ [3]; the finite presentation follows then from the existence of open fundamental sets and from the connectedness and simple connectedness of D [3; Satz 2]. Let Ω be a fundamental set and C the set of elements $c \in G_{\mathbf{Z}}$ for which $\Omega \cap \Omega \cdot c \neq \emptyset$. By our condition (ii), C is finite. Let now H be a finite subgroup of $G_{\mathbf{Z}}$. It belongs to a maximal compact subgroup of $G_{\mathbf{R}}$, hence has a fixed point $P \in D$; if $g \in G_{\mathbf{Z}}$ is such that $P \cdot g \in \Omega$, then $g^{-1} \cdot H \cdot g$ leaves a point of Ω fixed, hence belongs to C.)

(*b*) Let G be semi-simple, connected, and f a rational representation, defined over $\mathbf{Q}$, of G in a finite dimensional vector space V. Take coordinates in $V_{\mathbf{Q}}$ with respect to which $G_{\mathbf{Z}}$ is represented by integral matrices, which is always possible [7; 6.3], and let X be an orbit of G. Then if X is closed, the integral points of X form a finite number of orbits of $G_{\mathbf{Z}}$ [7; Theorem 6.9]; if the union of X and the origin is a cone, the primitive integral vectors contained in X form a finite number of orbits of $G_{\mathbf{Z}}$ [5].

Applications. (*a*) Let $G = \mathrm{SL}(n, \mathbf{C})$, the group of complex $n \times n$ matrices with determinant 1, f be the natural representation of G in the space of $n \times n$ symmetric matrices. The set X of symmetric matrices with a given non-zero determinant is then a closed orbit of G, and the first assertion shows that it contains only a finite number of classes of integral quadratic forms, a well known statement of Hermite [12].

(*b*) Let $G = 0(F)$ be the orthogonal group of the rational indefinite non-degenerate quadratic form F, and f be the identical representation of G. By Witt's theorem, the cone of isotropic vectors minus the origin is an orbit of G. The second assertion implies therefore that the primitive integral isotropic vectors form a finite number of orbits for the group of units of F, which is also well-known from reduction theory.

If G is an algebraic matric group defined over a number field k, then its group of units is by definition G_B, where B is the ring of algebraic integers of k. This case can be reduced to the one considered here ($k = \mathbf{Q}$, $B = \mathbf{Z}$) by the so-called "restriction of the scalars" [31; § 1], and *mutatis mutandis*, the previous results also apply to these apparently more general groups of units [7, § 12]. Clearly, they are also valid for subgroups of finite index of groups of units.

6. Adele groups

So far we have essentially considered $G_{\mathbf{R}}$, $G_{\mathbf{Z}}$ and $G_{\mathbf{Q}}$. Now $G_{\mathbf{R}}$ is attached to G and to the completion $\mathbf{R}$ of $\mathbf{Q}$ with respect to the usual absolute value. Similarly, we may associate a group G_p to G to the p-adic valuation of $\mathbf{Q}$, where p is any prime. G_p is then the group of $n \times n$ invertible matrices with coefficients in the field Q_p of p-adic numbers, which satisfy the same equations as the elements of G; this makes sense, since these are polynomial equations with coefficients in $\mathbf{Q}$, hence in $\mathbf{Q}_p$. Endowed with the p-adic topology, G_p is a locally compact, totally disconnected group. The subgroup $G_{\mathbf{Z}_p}$ of elements of G_p with coefficients in the ring $\mathbf{Z}_p$ of p-adic integers and whose determinant is a p-adic unit is then compact and open. The systematic consideration of the p-adic algebraic groups G_p besides the real algebraic groups $G_{\mathbf{R}}$ has considerably widened the scope of the theory of linear algebraic groups, and I shall try to illustrate this with some examples.

Following a procedure familiar in number theory, we may associate to G a certain subgroup of the direct product of $G_{\mathbf{R}}$ by the G_p's, the *adele group* attached to G, to be denoted by G_A. By definition, an element x of this direct product is an adele of G if and only if for all but a finite number of p's, the component of x in G_p belongs actually to $G_{\mathbf{Z}_p}$. Thus G_A contains in particular the product $G_A^\infty = G_{\mathbf{R}} \times \Pi_p\, G_{\mathbf{Z}_p}$ of the locally compact group $G_{\mathbf{R}}$ by the compact groups $G_{\mathbf{Z}_p}$; the group G_A^∞ is therefore locally compact for the product topology. G_A itself admits one structure of topological, locally compact group in which G_A^∞ is an open subgroup.

If we apply this construction to the cases where G is isomorphic over $\mathbf{Q}$ to the additive group of $\mathbf{C}$, or to the multiplicative group of non-zero complex numbers, we get respectively the adele and idele groups of $\mathbf{Q}$, in the sense of Chevalley.

The group $G_{\mathbf{Q}}$ is contained in $G_{\mathbf{R}}$ and in all G_p's, and any $a \in G_{\mathbf{Q}}$ belongs to $G_{\mathbf{Z}_p}$ for all but a finite number of p's, those which occur in the denominators of the coefficients or in the determinant of x; therefore, the element of the direct product $G_{\mathbf{R}} \times \Pi_p\, G_p$ all of whose components are equal to a is an adele of G, called a *principal adele*. Thus we may (and shall) identify $G_{\mathbf{Q}}$ with a subgroup of G_A, so to say diagonally embedded, which is easily seen to be discrete.

Adele groups were introduced by Ono [19, 20], Tamagawa, and also by M. Kneser in the case of orthogonal groups. They can also be associated to groups defined over finite number fields, or over function fields of one variable with finite constant field [31]. As above, the number field case can be reduced to the one considered here by restriction of the scalars [31].

7. Finiteness theorems for adele groups

The definition of G_A puts on almost equal footing the p-adic completions of $\mathbf{Q}$ and $\mathbf{R}$ and, roughly speaking, the pair $G_{\mathbf{R}}, G_{\mathbf{Z}}$ appears as the localization, at the infinite prime spot, of the pair $G_A, G_{\mathbf{Q}}$. However, this symmetry is not a true one, the infinite prime spot plays a special role among the primes, and the results discussed in the previous sections imply rather easily some global properties of adele groups [4, 6], proved for solvable groups first by Ono [20], in particular:

(*a*) If G is connected, the number of double cosets $G_A^\infty \cdot x \cdot G_{\mathbf{Q}}$ of G_A is finite.

If G is the proper orthogonal group of a non-degenerate rational quadratic form F, these double cosets correspond to the proper classes in the genus of F. If G is obtained by restriction of the scalars from the multiplicative group of a number field k, they correspond to the ideal classes of k.

(*b*) There exist closed or open fundamental sets for $G_{\mathbf{Q}}$ in G_A, which have finite Haar measure when G is semi-simple or, more generally, when the identity component of G has no non-trivial rational character defined over $\mathbf{Q}$.

A fundamental set is here, in analogy with the definition given in § 2, a subset of G_A which meets every coset $x \cdot G_{\mathbf{Q}}$ ($x \in G_A$) and intersects only a finite number of its right translates by elements of $G_{\mathbf{Q}}$. Under the last condition of (*b*), every left invariant Haar measure on G_A is also right invariant (i.e. G_A is unimodular) and defines an invariant measure on $G_A/G_{\mathbf{Q}}$; our assertion implies in this case that $G_A/G_{\mathbf{Q}}$ has finite volume for any such measure.

(*c*) $G_A/G_{\mathbf{Q}}$ is compact if G is semi-simple and $G_{\mathbf{Q}}$ consists of semi-simple elements.

The condition for compactness is the same as for $G_{\mathbf{R}}/G_{\mathbf{Z}}$. Another proof of (*c*) is given in [33].

8. Tamagawa numbers

Let G be connected, unimodular, and ω a left-invariant rational differential m-form ($m = \dim G$) defined over $\mathbf{Q}$. Any other such form is a rational multiple of ω. The latter defines in a natural way Haar measures ω_∞ and ω_p on $G_{\mathbf{R}}$ and G_p (p any prime). Assume that the product of the volumes v_p of the groups $G_{\mathbf{Z}_p}$ for those measures converges absolutely. Then the product measure of ω_∞ and the ω_p's is defined on G_A^∞ and gives rise to a Haar measure ω_A on G_A, which does not change if ω is a replaced by a rational multiple of itself. Weil has proposed to call the volume $\tau(G)$ of $G_A/G_{\mathbf{Q}}$ with respect to this normalized Haar measure the *Tamagawa number of* G. The above convergence condition is fulfilled when G is semisimple (3); if it is not, we may still multiply the ω_p's by suitable convergence factors λ_p and define the Tamagawa number of G with respect to the λ_p's (for all this, see [31, Chapter 2]).

It was noticed by T. Tamagawa and M. Kneser that when G is the proper orthogonal group of a non-degenerate rational quadratic form F in at least three variables, the equality $\tau(G) = 2$ is essentially equivalent to some of the main results obtained by Siegel in his famous series of papers on the analytic theory of quadratic forms [26].(4) Briefly, this comes about in the following way: it follows in a straightforward fashion from the definitions, for any semi-simple group, that $\tau(G) = v_\infty \cdot \Pi_p v_p$, where v_p (p prime) is as

(3) This was noticed by Serre some years ago. It follows from the relation between v_p and the number of rational points of the reduction mod p of G when the reduction is "good" [31] and from the formulas giving the number of points of a simple algebraic group defined over a finite field.

(4) It corresponds to the case $m = n$, $\mathcal{S} = \mathcal{J}$ of Siegel (representations of a form into itself). As to the passage from there to the general case, see M. Kneser (*Math. Zeit.* 77 (1961), 188–194), A. Weil, (*Colloque sur la Théorie des Groups Algébriques*, Bruxelles, 1962).

above and where v_∞ is the sum of the volumes of $G_{\mathbf{R}}/G_{\mathbf{Z}}$ and of finitely many quotients $G_{\mathbf{R}}/\Gamma$, where Γ runs through groups of units for certain lattices which correspond to representatives of the double cosets $G_A^\infty \cdot x \cdot G_{\mathbf{Q}}$. If $\tau(G)$ is then equal to two, or to some simple more or less universal constant, which, so to say, comes from the outside, this means that $1/v_\infty$ may be expressed as an infinite product of factors v_p, where v_p depends only on G_p. In the orthogonal group case, v_∞ is Siegel's generalization of the measure of the genus of a positive form and v_p may be also interpreted as a p-adic density of numbers of units of F modulo high powers of p, whence the classical formulation of Siegel's theorem.

Similarly, if G is the symplectic group $\mathbf{Sp}(2n)$, then v_∞ is the volume of $G_{\mathbf{R}}/G_{\mathbf{Z}}$ and the equality $\tau(G)=1$ is equivalent to a formula of Siegel [28] expressing the volume of the fundamental domain for the modular group as an infinite product.

Thus the Tamagawa number has considerable interest. It has been computed in various cases by Tamagawa, Weil, Demazure [31], Ono [21, 34], but our knowledge of it is still very incomplete. In a number of cases, it is equal to the order of the fundamental group of G. However, Ono [34] has given an example of a connected semi-simple group whose Tamagawa number is not integral, although still rational. It is not known whether $\tau(G)$ is always a rational number. It is conjectured to be equal to one when G is connected and simply connected.

The known computations of Tamagawa numbers use induction, residues of certain analogues of zeta-functions, and the Poisson summation formula. This requires a mixture of additive and multiplicative features, and so far the method could be applied mainly to groups of invertible elements in certain algebras defined over $\mathbf{Q}$.

As to adele groups defined over function fields, Tamagawa numbers have been computed and analogues of (*a*), (*b*), (*c*) in § 7 have been proved only for certain classical groups [31]. For the finiteness theorems, the methods of [4, 6] are of no avail, in view of the lack of archimedean primes. A further difference is the need of a theory of algebraic groups defined over non-perfect fields. However the situation there might improve soon, since Grothendieck has recently obtained results of a "foundational" character on semi-simple groups defined over arbitrary fields, in fact on schemes of semi-simple groups, some of which are already used in [30].

9. Principal homogeneous spaces

Another problem involving p-adic algebraic groups as well as real algebraic groups and of considerable interest for arithmetic is the existence of rational points in principal homogeneous spaces.

A principal homogeneous space is so to say what remains of a group when you disregard the identity. A simple example is the affine space, as compared to the additive group of a vector space. More precisely, in the context of algebraic geometry, a principal homogeneous space is an algebraic variety V on which an algebraic group G acts, say on the right, by means of a map $G \times V \to V$ which is regular in the sense of algebraic geometry, so that for every $v \in V$, the map $g \to v \cdot g$ is a birational biregular map of G onto V. Here, contrary to our general convention, G need not be linear, defined over $\mathbf{Q}$,

and may be any group variety. The principal homogeneous space is said to be defined over a field k if k is a field of definition for G, V and the action of G onto V. Of course, once we have picked up a point $v \in V$, we may identify V with G acting on itself by means of right translations, but an important question is precisely over which fields we can choose rational points.

For example, let $\mathfrak{g}, \mathfrak{g}'$ be two Lie algebras over k which become isomorphic after extension of the groundfield to an algebraic closure $\bar{k}$ of k. Then the set of isomorphisms of $\mathfrak{g} \otimes \bar{k}$ onto $\mathfrak{g}' \otimes \bar{k}$ is in an obvious way a principal homogeneous space for the group of automorphisms of $\mathfrak{g}' \otimes \bar{k}$. It is defined over k, and if it has a rational point over an extension k' of k, this means that $\mathfrak{g} \otimes k'$ is isomorphic to $\mathfrak{g}' \otimes k'$. Principal homogeneous spaces occur in this fashion when studying isomorphism classes of geometric objects whose automorphism group is algebraic.

There is an obvious notion of isomorphism over k for principal homogeneous spaces of a group G defined over k. These isomorphism classes are in 1–1 correspondence with the elements of a Galois cohomology set, usually denoted $H^1(k, G)$. This is the first cohomology set of the Galois group of the separable closure k_s of k, with coefficients in G_{k_s}. It G is commutative, $H^1(k, G)$ is a commutative group; otherwise it is just a set with a zero element, the isomorphism class of principal homogeneous spaces which "split over k", i.e. which have rational points over k and are therefore isomorphic over k to G acting on itself by right translations.

Coming now to the case where $k = \mathbf{Q}$, a particularly interesting question is whether two geometric objects defined over $\mathbf{Q}$ which are isomorphic over every completion of $\mathbf{Q}$ are already isomorphic over $\mathbf{Q}$. That for instance a statement of this kind is true for rational quadratic forms is a classical theorem of Hasse. In our present set up, the question is whether a principal homogeneous space defined over $\mathbf{Q}$ and having rational points for all completions of $\mathbf{Q}$ also has a point rational over $\mathbf{Q}$; if not, what can be said about the number, call it $\nu(G)$, of classes of such principal homogeneous spaces? Hasse's theorem implies that $\nu(G) = 1$ when G is an orthogonal group. It can be shown that $\nu(G)$ is *finite for any linear algebraic group* G (see [4, 6] when G is reductive, connected, [8] for the extension to the general case). It has been known for a long time that $\nu(G)$ may be $\neq 1$ when G is an algebraic torus, or is not connected; recently Serre has provided an example where G is semi-simple and connected (yet unpublished). It is conjectured that $\nu(G) = 1$ when G is connected and simply connected; but this would not contain Hasse's theorem, and the situation is not well understood at all. Another conjecture relating local and global properties is that, G being simple over $\mathbf{Q}$, if $G_{\mathbf{R}}$ and all G_p's contain unipotent elements $\neq e$, then so does $G_{\mathbf{Q}}$. This would generalize the fact, (due to Hasse, too), that a rational quadratic form which represents zero rationally over every completion of $\mathbf{Q}$ also does so over $\mathbf{Q}$.

The two quoted theorems of Hasse on quadratic forms play a role in the computation of Tamagawa numbers, and it is somewhat tempting, at least to me, to think that there is some connection between the failure of $\nu(G)$ to be one and the failure of $\tau(G)$ to be an integer; but there is so far little evidence to back this up.

In our previous example with Lie algebras, the finiteness of $\nu(G)$ implies that the isomorphism classes of Lie algebras over $\mathbf{Q}$ which are isomorphic to

a given Lie algebra after extension of the groundfield to all completions of $\mathbf{Q}$ are finite in number. An example of Landherr [16] shows that there may be more than one class. That there is only one such class for Lie algebras of type G_2 or F_4 [1] is positive evidence for the above conjecture on simply connected groups.

The problem discussed so far is a global one, or rather relates local and global properties, and it is natural to complete it by local questions, over a p-adic field k (which is assumed here to be of characteristic zero). It is known that $H^1(k, G)$ is finite when G is linear, defined over k; moreover, the number of forms of G, i.e. the number of isomorphism classes over k of algebraic groups defined over k and isomorphic to G over an algebraic closure of k, is also finite [8]; the latter finiteness statement is not a special case of the former one, since the group of automorphisms of an algebraic group, unlike that of a Lie algebra, is not always an algebraic group. It is conjectured, and has been checked in many cases, that $H^1(k, G) = 0$ when G is connected and simply connected.

There are a number of partial results and conjectures on principal homogeneous spaces for linear groups, for which we refer to [14, 25]. The problems on principal homogeneous spaces are often meaningful for general group varieties, not necessarily matric groups, and in fact have attracted much attention in the case of abelian varieties (see e.g. [17, 29]).

10. Groups of S-units

When G is the group obtained by restriction of the scalars from the multiplicative group of a finite number field k, $G_{\mathbf{Z}}$ is nothing but the group of units of k, and its finite generation is part of the Dirichlet unit theorem. It is well known that this was generalized by Hasse and Chevalley to the group of S-units of k, where S is a finite set of primes, i.e. to the group of elements of k^* which are p-adic units for all primes p not in S. If now G is as before an algebraic matric group defined over $\mathbf{Q}$, we may similarly introduce the group $G_{\mathbf{Z}_S}$ of S-units of G. These are the elements of $G_{\mathbf{Q}}$ whose determinant is, up to sign, a power product of primes of S, and whose coefficients have denominators which are products of elements of S. If S is the empty set, then $G_{\mathbf{Z}_S} = G_{\mathbf{Z}}$.

M. Kneser has shown that if G_p has a compact set of generators for every $p \in S$, then $G_{\mathbf{Z}_S}$ is finitely generated. The proof uses the finite generation of $G_{\mathbf{Z}}$ and (a) of § 7. The condition on the G_p's is fulfilled, for any p, when G is reductive (i.e. fully reducible or, equivalently, isogeneous to the product of a semi-simple group by an algebraic torus), as follows from the generalized Bruhat lemma of [5, 30]. Earlier, H. Behr [3] had shown for certain classical groups that $G_{\mathbf{Z}_S}$ is finitely generated and finitely presented. To my knowledge it is not proved, although of course very likely, that the latter property is true for any reductive group G.

It is shown in [7] that if $f: G \to G'$ is an isogeny (a rational surjective homorphism with finite kernel) defined over $\mathbf{Q}$ of G onto an algebraic matric group G' also defined over $\mathbf{Q}$, then $f(G_{\mathbf{Z}})$ is commensurable with $G'_{\mathbf{Z}}$. Honda [13] has proved the analogous statement for $G_{\mathbf{Z}_S}$ when S is big enough (a restriction which Serre has recently removed (unpublished), see also [6]), and has also considered similar questions for abelian varieties.

Let us mention finally a problem which, as far as I know, is unsolved in the general case for groups of units (and *a fortiori* for groups of S-units): whether the commutator subgroup of $G_{\mathbf{Z}}$ has finite index in $G_{\mathbf{Z}}$ when G is semi-simple. It follows from results of Matsushima [18] that this is true when D (notation of § 1) is a product of irreducible bounded symmetric domains none of which is isomorphic to a unit ball and $D/G_{\mathbf{Z}}$ is compact. According to Steinberg, it is also true when G is the adjoint group of a semi-simple Lie algebra in Chevalley's normal form (unpublished).

11. Compact subgroups of p-adic groups

All this has aroused much interest in p-adic Lie groups and, to round off this report, we would like to mention some unsolved problems on them. An algebraic matric p-adic group is at the same time a p-adic Lie group (p-adic analytic manifold endowed with a group structure for which product and inverse are analytic functions). From the two approaches to Lie group theory mentioned at the beginning of the introduction, the second one has yielded results valid for p-adic as well as for real groups and has been so far more successful than the first one. In particular our knowledge of maximal compact subgroups is still scanty. Although they do not always exist (e.g. in the additive group of a p-adic field), it seems rather likely that a semi-simple algebraic matric group L always has some; this has been checked by Bruhat [9, 10], Satake for many classical groups and groups in Chevalley's normal form(5). However, they may form several conjugacy classes [10].

Let B be a maximal irreducible subgroup of L which can be put in triangular form An important question is to know whether there exists a compact subgroup K of L such that

$$L = K \cdot B, \tag{1}$$

$$L = K \cdot A \cdot K, \tag{2}$$

where A is a Cartan subgroup of B. In the real case, (1) is the Iwasawa decomposition, and (2) follows from a well-known theorem of E. Cartan. (1) and (2) have been checked for many classical groups by Bruhat [9], Satake, and play a basic role in the study of infinite dimensional representations and of zonal functions. In general, it is only known that L is the union of a finite number of double cosets $K \cdot x \cdot B$ for any compact and open subgroup K.

12. Concluding remarks

I hope that this report, although not exhaustive,(6) has given an idea of the general trends and aims of these investigations, and of some of their connections with arithmetic. The general results obtained so far are chiefly finiteness theorems, that is of a qualitative nature. They have yet to be matched by quantitative results of similar scope, e.g. on Tamagawa numbers, or on the true extent of the generalization of Hasse's theorem. One

(5) (Added in proof): A general proof has been recently indicated to me by R. P. Langlands.

(6) In particular, we refer to [14] for approximation problems in adele groups.

possible obstacle may be the lack of a good device to make use of class field theory, or rather of some consequences of it which occur frequently when dealing with classical groups. But, more likely, what is required first of all are new developments of an analytic nature, notably on Eisenstein series, automorphic functions, representation theory, and some of this not only for real algebraic groups, but also for p-adic and adele groups. No doubt this is a vast program, whose completion does not seem in sight at present. However these questions have recently been and are being studied in several quarters from different angles. There are substantial results and promising lines of attack. It is therefore not too unlikely that further progress will be achieved in the not too distant future and that the arithmetic properties of algebraic groups surveyed here will become part of a general arithmetic and analytic theory of algebraic groups.

REFERENCES

[1]. ALBERT, A. A. & JACOBSON, N., On reduced exceptional simple Jordan algebras. *Ann. Math.*, 66 (1957), 400–417.

[2]. ANDREOTTI, A. & GRAUERT, H., Algebraische Körper von automorphen Funktionen. *Nachr. Akad. d. Wiss. Göttingen, math.-phys. Klasse 1961*, Heft 3, 40–48.

[3]. BEHR, H., Über die endliche Erzeugbarkeit verallgemeinerter Einheitsgruppen. (To appear.)

[4]. BOREL, A., Some properties of the adele groups attached to algebraic groups. *Bull. Amer. Math. Soc.*, 67 (1961), 583–585.

[5]. —— Ensembles fondamentaux pour les groupes arithmétiques. *Coll. Théorie des Groupes Algébriques*. Bruxelles, 1962.

[6]. —— Finiteness theorems for adele groups over number fields. (To appear.)

[7]. BOREL, A. & HARISH-CHANDRA, Arithmetic subgroups of algebraic groups. *Ann. Math.*, 75 (1962), 485–535.

[8]. BOREL, A. & SERRE, J.-P., Théorèmes de finitude en cohomologie galoisienne. (En préparation.)

[9]. BRUHAT, F., Sur les représentations des groupes classiques p-adiques I, II, *Amer. J. Math.* 83 (1961), 321–338, 343–368.

[10]. —— Sur les sous-groupes compact maximaux des groupes algébriques p-adiques. *Coll. Théorie des Groupes Algébriques*. Bruxelles, 1962.

[11]. GEL'FAND, I. M., Automorphic functions and theory of representations. *Proc. Internat. Congress Math. Stockholm*, 1962. Uppsala 1963.

[12]. HERMITE, C., *Œuvres Complètes*, Vol. 1. Gauthier-Villars, Paris, 1905.

[13]. HONDA, K., Isogenies, rational points and section points of group varieties. *Japan J. Math.*, 30 (1960), 84–101.

[14]. KNESER, M., Einfach zusammenhängende algebraische Gruppen in der Arithmetik. *Proc. Internat. Congress Math. Stockholm*, 1962. Uppsala 1963.

[15]. KORKINE, A. & ZOLOTAREFF, G., Sur les formes quadratiques. *Math. Ann.*, 6 (1873), 366–389.

[16]. LANDHERR, W., Liesche Ringe vom Typus A über einem algebraischen Zahlkörper (die lineare Gruppe) und hermitesche Formen über einem Schiefkörper. *Abh. Math. Sem. Hamburg*, 12 (1938), 200–241.

[17]. LANG, S., Some theorems and conjectures in diophantine equations. *Bull. Amer. Math. Soc.*, 66 (1960), 240–249.

[18]. MATSUSHIMA, Y., On the first Betti number of compact quotient spaces of higher dimensional symmetric spaces, *Ann. Math.*, 75 (1962), 312–330.

[19]. ONO, T., Sur une propriété arithmétique des groupes commutatifs. *Bull. Soc. Math. France*, 85 (1957), 307–323.

[20]. —— On some arithmetic properties of linear groups. *Ann. Math.*, 70 (1959), 266–290.

[21]. —— Arithmetic of algebraic tori, ibid. 74 (1961), 101–139.

[22]. PYATETSKI-ŠAPIRO, *Geometry of Classical Domains and Automorphic Functions*. Moscow, 1961.

[23]. SATAKE, I., On compactifications of the quotient spaces for arithmetically defined discontinuous groups. *Ann. Math.*, 72 (1960), 555–580.

[24]. SELBERG, A., Discontinuous groups and harmonic analysis. *Proc. Internat. Congress Math. Stockholm*, 1962, 177–189. Uppsala 1963.

[25]. SERRE, J-P., Cohomologie galoisienne des groupes algébriques linéaires. *Coll. Théorie des Groupes Algébriques*. Bruxelles, 1962.

[26]. SIEGEL, C. L., Über die analytische Theorie der quadratischen Formen. *Ann. Math.*, 36 (1935), 527–606; 37 (1936), 230–263; 38 (1937), 212–291.

[27]. —— Einheiten quadratischer Formen. *Abh. Math. Sem. Hamburg*. 13 (1939), 209–239.

[28]. —— Symplectic geometry. *Amer. J. Math.*, 65 (1943), 1–80.

[29]. TATE, J., *WC*-groups over p-adic fields. *Séminaire Bourbaki*, Exp. 157, Paris, Décembre 1957.

[30]. TITS, J., Groupes semi-simples isotropes, *Coll. Théorie des Groupes Algébriques*. Bruxelles, 1962.

[31]. WEIL, A., *Adeles and Algebraic Groups*. Notes by M. Demazure & T. Ono. The Institute for Advanced Study, Princeton, 1961.

(Added in proof):

[32]. BOREL, A., Compact Clifford-Klein forms of symmetric spaces. (To appear in *Topology*.)

[33]. MOSTOW, G. D. & TAMAGAWA, T., On the compactness of arithmetically defined homogeneous spaces, *Ann. Math.*, 76 (1962), 446–463.

[34]. ONO, T., On the Tamagawa number of algebraic tori. (To appear.)

LOGIC, ARITHMETIC, AND AUTOMATA(1)

By ALONZO CHURCH

This paper is a summary of recent work in the application of mathematical logic to finite automata, and especially of mathematical logic beyond propositional calculus.

To begin with a sketch of the history of the matter, let us recall that application of the "algebra of logic", i.e., elementary Boolean algebra, to the analysis of switching circuits was first suggested by Ehrenfest [A]. Nothing came of Ehrenfest's remark for many years, and it seems to have remained wholly unknown outside of Russia. Yanovskaya [G] says that details of the suggested application were worked out by Shestakoff in 1934–35. However, Shestakoff's candidate's dissertation, embodying the material, was presented to the University of Moscow in 1938, and the earliest publications by Shestakoff are [D] and [E] in 1941. Meanwhile the same idea had occurred independently to Nakasima and Hanzawa [B] and Shannon [C]. For some time the development of the idea proceeded independently in Russia, in Japan, and in the United States, the three lines of development having had at first no influence on one another.

This use of Boolean algebra is now widely familiar, and therefore requires no elaboration here. It is usually taken to be a Boolean algebra of cardinal number 2 that is used, although the character of the application would more naturally suggest propositional calculus. Use of the Boolean algebra and of propositional calculus are equivalent in a way that is well known. The choice of Boolean algebra is advantageous if algebraic methods and results are to be employed. But otherwise there is a certain artificiality in allowing only equations and inequalities to be asserted. And for further application of mathematical logic, the choice of propositional calculus provides a better basis.

Mathematical logic beyond propositional calculus is first applied to automata theory in the paper of McCulloch and Pitts [16], in which the context is biological. The authors are concerned with analyzing the behavior of a net of neurons and with the question of the existence of, and of finding, a neural net having some specified behavior. But their hypotheses about the behavior and the interaction of neurons are such that these questions become entirely similar to corresponding questions about electronic digital computing circuits. The relevance of the ideas of McCulloch and Pitts to the theory of digital computing circuits was noticed by John von Neumann, and it was evidently this that led him to suggest application of mathematical logic to the latter.

The only published reference to von Neumann's contribution to the matter is Hartree [13], where not only McCulloch and Pitts but also unpublished suggestions of both von Neumann and A. M. Turing are referred to. I am indebted to H. H. Goldstine for calling my attention to the privately cir-

(1) For support of the work represented by this paper the author is indebted to the National Science Foundation of the United States.

culated paper, von Neumann [18], in which there is some detailed discussion of the relationship between the McCulloch–Pitts neural nets and digital computing networks, but without use of mathematical logic.

The logical system employed by McCulloch and Pitts is first-order arithmetic. Both ordinary and bounded quantifiers are used, but not definition by recursion.

Hartree uses a singulary free-variable functional (or "predicate") calculus, having singulary predicates and numerical variables, but neither quantifiers nor recursions. There is a "delay operator" in the sense that not only a variable t but also $t+1$, $t+2$, etc., may appear as argument of a predicate.

The use of (in effect) such a singulary free-variable functional calculus also developed independently in Japan. The earliest paper that I have seen is Gotô [12]. But it is possible that some of the earlier papers to which Gotô refers may constitute at least a partial exception to the statement made above, that the first application of mathematical logic beyond propositional calculus was by McCulloch and Pitts.

The use of recursions in the treatment of circuits is found in Berkeley [1], and more explicitly in Burks and Wright [5]. The idea that a particular circuit or finite automaton can be completely characterized by giving a set of recursion equations—or in this context, since the functions "defined" by the recursions are propositional functions, we shall speak rather of recursion equivalences—appears first in the paper of Burks and Wright, but these authors do not have a recursion schema to which admissible recursions must conform.

The application of restricted recursive arithmetic to automata was introduced in Church [6] and [7]. This system may be described as obtained from singulary free-variable functional calculus by adding a rule of proof by mathematical induction and a rule of definition by recursion, as given below. It is therefore a specialized form of the recursive arithmetic of Skolem [19].

An abstract definition of finite automaton appears in Kleene [14], a paper of restricted circulation, and in Kleene [15], which is the published form of the same paper. The "logical nets" of Burks and Wright [5] may also be regarded as providing such a definition. But Kleene's definition is more satisfactory, and we shall here use it in the following slightly modified and generalized form.

A *finite automaton* consists of a finite number of *elements*, each of which is capable of a finite number, $\leqslant n$, of different *states*. Time is measured in discrete *instants*, $t=0,1,2,\ldots$, beginning at an *initial instant* and extending indefinitely. The elements are distinguished as *input elements*, *intermediate elements*, and *output elements*. The states of the input elements, at any instant, constitute the *input state*, at that instant; the states of the output elements, at any instant, similarly constitute the *output state*; and the states of the intermediate and output elements constitute the *internal state*. Except the input elements, whose states are imposed from outside, the state of any element at a given instant is completely determined, in some noncircular way, by (1) the states of certain other elements at the given instant, and (2) the states of certain elements, itself and others, at instants which precede the given instant by at least 1 and by not more than a certain fixed maximum *span* $h+1$.

(More fully, the foregoing statement is a definition of "discrete synchronous deterministic finite automaton with outputs." But the longer phrase

is for the sake of distinction from a number of other notions with which we are not here concerned, and let us therefore say simply "finite automaton" or "automaton".)

Though there is a certain loss of structure in doing so, it is quite usual to take $n=2$, $h=0$. In this lecture we shall take $n=2$, in order to identify the two states of an element with the two truth-values, T and F. But we preserve a general value of h, $\geqslant -1$.

It is also possible to abstract from the elements of the automaton and to consider only certain numbers called input states, output states, and internal states, and the way in which the successive output states and internal states are determined by the input states. Namely there are to be a finite number each, $\geqslant 1$, of input states, output states, and internal states; the output state at any instant is to be uniquely determined by the internal state at the same instant; and the internal state at any instant is to be uniquely determined by (1) the input state at that instant and (2) the input states and the internal states at instants preceding the given instant by not more than $h+1$. In this case we have a *finite transition system.*

In either case, automaton or transition system, the sequence of input states from $t=0$ to $t=t_0$ is called an *input sequence*, and the sequence of output states from $t=0$ to $t=t_0$ is called an *output sequence*. The function which, for an input sequence from $t=0$ to $t=t_0$, as argument, has as value the output state at $t=t_0$ is the *behavior function*. The null sequence is also to be counted as an input sequence, but one for which there is no corresponding output state and hence no value of the behavior function.

Two automata, or two transition systems, are said to be *equivalent* if they have the same behavior function. Evidently, two automata are equivalent if and only if their transition systems are equivalent. In particular, two automata with the same transition system are equivalent—but they do not necessarily have the same *structure*, in an obvious sense of that word.

A class of input sequences is an *event*.

An event is *represented* by an automaton if the set of values of the behavior function for input sequences that belong to the event is disjoint from the set of values of the behavior function for other input sequences.

The foregoing (together with the definition of regularity of an event, to be given below) are the essentials of the terminology of automata theory which we shall need. It should be added that the terminology in the field is not in a very settled state, and our terminology here is somewhat eclectic. We have followed Büchi [2] in adopting a terminology that maintains the distinction between automaton and transition system.

We turn now to statement of the primitive basis of restricted recursive arithmetic. And for expository convenience we adopt a formulation which emphasizes brevity of statement rather than economy of primitives.

The primitive symbols of restricted recursive arithmetic are as follows:

i. Numerical variables $t, x, y, z, \ldots$ (having the natural numbers as their range).

ii. The symbol 0 for the number zero.

iii. The accent $'$ as a notation for successor (of a natural number).

iv. Singulary primitive predicates $i_1, i_2, i_3, \ldots$. (The primitive predicates, and other singulary predicates introduced by recursion, are to be understood as denoting propositional functions of natural numbers.)

v. Parentheses ( ) as notation for application of a singulary propositional function to its argument.

vi. The letters T and F to denote the truth-values, truth and falsehood respectively.

vii. Connectives of propositional calculus (any sufficient set), and brackets [ ] for use in connection with them.

Then a *term* consists of the symbol 0, or the symbol 0 followed by any number of accents, or a numerical variable, or a numerical variable followed by any number of accents.

An *elementary formula* consists, either of one of the letters T or F, or of a singulary predicate followed by a term between parentheses.

Further *well-formed formulas* are constructed from the elementary formulas by means of connectives of propositional calculus. (We may sometimes abbreviate "well-formed formula" as "wff"; or we may say simply "formula" if well-formedness is clear from the context.)

The primitive rules of restricted recursive arithmetic are the five following:

1. To assert any formula that is tautologous according to laws of propositional calculus.

2. From $\mathrm{A} \supset \mathrm{B}$ (as major premiss) and A (as minor premiss) to infer B.

3. From any asserted formula to infer the result of substituting any term for any numerical variable throughout.

4. From $\mathrm{S}(t) \supset \mathrm{S}(t')$ (as major premiss) and $\mathrm{S}(0)$ (as minor premiss) to infer $\mathrm{S}(t)$.

5. A rule of "definition" (or better, introduction) by recursion, allowing us at any stage to introduce new predicates in accordance with the schema of wider restricted recursion as given below.

Rules 2, 3, and 4 are, in order, *the rule of modus ponens, the rule of substitution,* and *the rule of mathematical induction.*

In Rule 4, $\mathrm{S}(t)$ is not necessarily an elementary formula, but is any assertible wff; $\mathrm{S}(t')$ is the result of substituting t' for t throughout $\mathrm{S}(t)$; and $\mathrm{S}(0)$ is the result of substituting 0 for t throughout $\mathrm{S}(t)$.

The schema of "wider restricted recursion" referred to in Rule 5 is as follows:

$$\mathrm{r}_i(j) \equiv \mathrm{P}_{ij}[\mathrm{s}_1(0), \mathrm{s}_2(0), \ldots, \mathrm{s}_k(0), \mathrm{s}_1(1), \mathrm{s}_2(1), \ldots, \mathrm{s}_k(1), \ldots, \ldots, \mathrm{s}_1(j), \mathrm{s}_2(j), \ldots, \mathrm{s}_k(j)]$$

$$\mathrm{r}_i(t+h+1) \equiv \mathrm{Q}_i[\mathrm{s}_1(t), \mathrm{s}_2(t), \ldots, \mathrm{s}_k(t), \mathrm{s}_1(t+1), \mathrm{s}_2(t+1), \ldots, \mathrm{s}_k(t+1), \ldots, \ldots, \mathrm{s}_1(t+h+1), \mathrm{s}_2(t+h+1), \ldots, \mathrm{s}_k(t+h+1), \mathrm{r}_1(t), \mathrm{r}_2(t), \ldots, \mathrm{r}_n(t), \mathrm{r}_1(t+1), \mathrm{r}_2(t+1), \ldots, \mathrm{r}_n(t+1), \ldots, \ldots, \mathrm{r}_1(t+h), \mathrm{r}_2(t+h), \ldots, \mathrm{r}_n(t+h)]$$

Here $i=1,2,\ldots,n$ and $j=0,1,\ldots,h$, so that the schema consists of $(h+1)n$ equivalences altogether. The new predicates introduced by the recursion are $\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_n$, which may be any n new letters not previously used; and $\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_k$ are given predicates, which may be either primitive predicates or predicates which were introduced in previous recursions. For convenience in indicating the form of the equivalences, $t+1$ has been written in place of what should properly be written as t', $t+2$ in place of t'', and so on; thus, e.g., $t+h$ indicates a letter t with h accents after it. Similarly, 1 has been written in place of $0'$, 2 in place of $0''$, j in place of a 0 with j accents after it. Then finally the device has been adopted of indicating the elementary

parts of a wff, other than the elementary parts T and F, by writing them in brackets. Thus, e.g.,

$$\mathrm{P}_{12}[\mathrm{s}_1(0), \mathrm{s}_2(0), \ldots, \mathrm{s}_k(0), \mathrm{s}_1(1), \mathrm{s}_2(1), \ldots, \mathrm{s}_k(1), \mathrm{s}_1(2), \mathrm{s}_2(2), \ldots, \mathrm{s}_k(2)]$$

indicates a wff P_{12} that is built up by means of connectives of propositional calculus from some or all of the elementary formulas $\mathrm{s}_1(0), \mathrm{s}_2(0), \ldots, \mathrm{s}_k(0)$, $\mathrm{s}_1(0'), \mathrm{s}_2(0'), \ldots, \mathrm{s}_k(0'), \mathrm{s}_1(0''), \mathrm{s}_2(0''), \ldots, \mathrm{s}_k(0''), T, F$.

Now in order to represent the action of a given automaton by means of restricted recursive arithmetic, we may use a different predicate to correspond to each element of the automaton—primitive predicates $i_1, i_2, \ldots$ for the input elements, and other letters as convenient for the remaining elements. Normally we shall use $o_1, o_2, \ldots$ for the output elements.

If p is the predicate that corresponds in this way to the element e of the automaton, then $\mathrm{p}(t)$ is to denote the state of e at time t. Thus $\mathrm{p}(0)$ denotes T or F according as the state of the element e at time 0 is T or F, $\mathrm{p}(1)$ denotes T or F according as the state of the element e at time 1 is T or F, and so on. And $\mathrm{p}(t)$ may therefore be understood as asserting that the element e is in state T ("is operated," or the like) at time t.

The action of the automaton is then expressed by writing a set of recursion equivalences in the form of one or more wider restricted recursions, constituting definitions by recursion of the intermediate and output predicates (i.e., the predicates that correspond to intermediate and output elements), in terms of the input predicates $i_1, i_2, \ldots, i_\mu$. Given a particular automaton, and the specific rule by which the state of each element at any instant is determined from the states of that element at preceding instants and the states of other elements at the same and preceding instants, the corresponding set of recursion equivalences that characterizes the automaton is then uniquely determined. And conversely, given a set of recursion equivalences of the kind just described, there is then a unique corresponding automaton.

The one-to-oneness of the correspondence between automata and sets of recursion equivalences has been secured by adopting a restricted form of recursion schema. This does not mean any restriction upon the kinds of elements that are available—instead there is an assumption that elements of arbitrary kind are available, within the general definition of automaton as given above, and if we wish a limitation to elements of special kind (e.g. to "and" elements, "or" elements, "not" elements, and one-unit delay elements), it is necessary to impose further restrictions on the recursion schema. But the restrictions which are involved in the schema of wider restricted recursion, as compared to the recursion schema which would be natural if we did not have the application to automata in mind, are dictated rather by two general principles of causality which may be stated briefly as follows: (I) No direct causal action of the future upon the present. (II) No direct causal action of the remote past upon the present. The question, how far back into the past is to be regarded as the immediate rather than the remote past, has an answer in the span $h+1$, which serves as a measure of this.

Two sets of recursion equivalences of the kind here in question are called *equivalent* if the corresponding automata are equivalent. Evidently this notion may have a direct definition, in terms of the recursion equivalences

themselves rather than the corresponding automata, provided that some subset of the predicates, not including any of the input predicates $i_1, i_2, \ldots, i_u$ is designated as output predicates. Hence we may state:

THE SIMPLIFICATION PROBLEM. Given a set of recursion equivalences of the kind here in question, to find one equivalent to it, which has a specified form, corresponding to a limitation to elements of special kind from which the automaton is to be constructed, and which is simplest in some suitable sense of simplicity.

The definition of simplicity adopted will evidently depend on the kinds of elements which are regarded as available to construct the automaton, and perhaps on some weighting of them. Thus there is properly not one problem but many. But little has been done with the simplification problem for automata, except in the case of combinational circuits, i.e., automata for which the span $h+1$ is 0, where only propositional calculus is required for the treatment.

In the case of transition systems it is natural to characterize simplicity as minimizing the number of internal states. This has often been followed. But it is clear that such minimization of the number of internal states will be at most a necessary condition of simplicity of a corresponding automaton, and the relevance to the simplification problem for automata is therefore uncertain.

Both the decision problem and the synthesis problem for automata involve a condition which the automaton is required to satisfy, and which we shall call the *synthesis requirement*. The synthesis requirement must be stated in some formalized language L, and we shall in this context suppose that L is either restricted recursive arithmetic itself or an extension of restricted recursive arithmetic obtained by adjoining additional notations—such as, e.g., the binary predicate $=$, the sign $+$ of addition, the quantifiers—together with rules for them. Then we may state the two problems as follows:

THE SYNTHESIS PROBLEM. Given as synthesis requirement a wff of L, containing the input predicates $i_1, i_2, \ldots, i_\mu$ and the output predicates $o_1, o_2, \ldots, o_\nu$, to find definitions by recursion of $o_1, o_2, \ldots, o_\nu$, and any number of intermediate predicates, from $i_1, i_2, \ldots, i_\mu$, in the form of one or more wider restricted recursions, such that the synthesis requirement is satisfied independently of what particular propositional functions are denoted by $i_1, i_2, \ldots, i_\mu$. Or if the given synthesis requirement is impossible in the sense that there is no automaton, and no set of recursion equivalences that satisfies it, this fact is to be ascertained.

THE DECISION PROBLEM. Given as synthesis requirement a wff of L, as before, and given a proposed solution of the synthesis problem in the form of a set of recursion equivalences (by wider restricted recursion), to determine whether the synthesis requirement is in fact a consequence of the recursion equivalences.

Both the synthesis problem and the decision problem may be divided into cases, according to the form of the synthesis requirement, or the language L in which it is expressed. Solution of any case of the synthesis problem implies, *a fortiori*, solution of the corresponding case of the decision problem, because a decision problem may always be turned into a synthesis problem by including the given recursion equivalences as a part of the synthesis requirement.

Existing results in regard to the synthesis problem and the decision

problem are summarized in the following table. But it should be emphasized that the indicated synthesis algorithms and decision algorithms are by no means always of practicable length.

	Synthesis requirement expressed in restricted recursive arithmetic	Synthesis problem	Decision problem
Case 1	with one numerical variable only	Solved in [7]	Solved in [11]
Case 2	with any number of numerical variables	Solved in [7]	Solvable by the method of [11]
Case 3	plus quantifiers	Partial solution in [7]	Solved in [7]
Case 4	plus = and <	Incomplete sketch of solution in [7], properly an open problem	Solution is a consequence of [23].
Case 5	plus +, =, and quantifiers	Proved unsolvable [4]	Open

Case 1 of the decision problem was solved by Joyce Friedman [11], the earliest result. The problem of determining the equivalence of two automata which are both given by means of recursion equivalences in the form of wider restricted recursions can be formulated as a subcase of Case 1 of the decision problem, and is therefore covered by Miss Friedman's solution. Her method is also immediately applicable to solution of Case 2 of the decision problem, as pointed out by McNaughton in his review.

Case 1 of the synthesis problem was first solved by me. A modification of the synthesis algorithm, by which it is substantially shortened, was then suggested by J. B. Wright, and this is incorporated in the solution as given in [7]. A subcase of this case of the synthesis problem was also solved independently by Wang [22], without recognizing the relationship to recursive arithmetic.

The unsolvability of Case 5 of the synthesis problem is, of course, in the usual sense, that there is no uniform effective procedure which can be given once for all and by which alone every instance of this case can be solved.

A variety of cases of the synthesis problem and of the decision problem, intermediate between Case 2 and Case 5, immediately suggest themselves as being still open. We may, for example, adjoin the sign of addition, +, to restricted recursive arithmetic. Or we may adjoin the sign + and the binary predicate = to restricted recursive arithmetic; or, as would be equivalent to this, we may adjoin the ternary predicate Σ, as sign of the ternary relation of addition. Other possibilities are restricted recursive arithmetic plus + and quantifiers; or plus = and quantifiers; or plus < and quantifiers. Or, as suggested by Büchi [2], we may consider adjoining singulary predicate-variables and quantifiers for them.

If to restricted recursive arithmetic we adjoin a sign of addition and a sign of multiplication and the sign = (of equality) and quantifiers, we obtain a formalized language which includes first-order arithmetic as a part. It follows that the decision problem (*Entscheidungsproblem*), and hence *a fortiori* the synthesis problem, is then unsolvable—cf. [4].

Turning now to matters less closely related to recursive arithmetic, we introduce the definition of "regularity of an event" in the sense of Kleene, employing for this purpose the "regular-expression language" of McNaughton and Yamada [17]. We follow the original notation of McNaughton and Yamada, except that we change their $\sim$ to a $-$, out of reluctance to confuse notations of propositional calculus and of class calculus.

We suppose that there are $m+1$ input states, of an automaton under consideration, and let them be represented by the numbers $0, 1, \ldots, m$. We use (k) as a notation for the unit class of the input sequence which consists of the single input state k. Or, when convenient, and especially if $m \leqslant 9$, the parentheses enclosing the numeral k may be omitted as an abbreviation.

The notations $\cup$, $\cap$, $-$, and Λ are employed with their usual set-theoretic meanings, i.e., for union of classes, intersection of classes, complement, and null class respectively. The notation ϕ is employed for the unit class of the null sequence.

If a and b are events, then $a \cdot b$ (abbreviated when convenient as ab) is the event whose members are every sequence that is composed of the concatenation of a sequence belonging to a followed by a sequence belonging to b.

If a is an event, then the event a^* is the smallest class of sequences which contains (as members) the null sequence and all sequences belonging to a and is closed under concatenation.

The regular expressions are: (k), for any suitable numeral k; and Λ; and ϕ; and any expression built up out of these by means of $\cup$, $\cap$, $-$, $\cdot$, and $*$ (with, of course, suitable use of brackets).

The regular expressions are intended as names of events, in a way that is implicit in the foregoing explanation. And we may then define an event as being regular if, under some choice of m, there is a regular expression which denotes it. We have the important result:

KLEENE'S THEOREM. *An event can be represented by a finite automaton if and only if it is regular.*

This was first proved by Kleene in [14] and [15], under a definition of regularity equivalent to the one we have just given.

Since the behavior function of an automaton with one binary (i.e., two-state) output is uniquely characterized by giving the event for which the output is T, McNaughton and Yamada propose their regular-expression language for practical use in specifying the behavior of such an automaton, and recommend the redundancy of its notation—as compared, e.g., to the more economical regular-expression languages of [15] and [8]—as facilitating this. For example, if we assume two input states, 0 and 1, the expression

$$-[(0 \cup 1)^*000(0 \cup 1)^*] \cup (0 \cup 1)^*111 - [(0 \cup 1)^*000(0 \cup 1)^*]$$

specifies an automaton as having output T at time t if and only if either there have never been three consecutive 0's in the input sequence up to and including the time t (this accounts for the part of the expression up to the end of the first square bracket) *or* (this accounts for the next symbol, $\cup$, in the expression) there have been three consecutive 1's in the input sequence since the last three consecutive 0's (this accounts for the remainder of the expression).

Kleene proposed the problem of a characterization of regular events directly in terms of their expression in a formalized language of ordinary kind, such as one of the usual formulations of first- or second-order arithmetic. This problem has since been solved by Trachtenbrot [20], and in a different way by Büchi [2] and Elgot [10].

These results of Büchi and Elgot are closely related to one another, and the substance of them was originally announced in the joint abstract [3]. Specifically, representability of an event by a finite automaton is equivalent to expressibility in any one of the three following formalized languages:

(1) *Weak second-order arithmetic* (Büchi). The primitive symbols are: numerical variables; singulary predicate-variables; the notation ( ) for application of a propositional function to its argument; the symbol 0 for the number 0; the accent $'$ as a notation for successor; T, F, and connectives of the propositional calculus; quantifiers on numerical variables; quantifiers on singulary predicate-variables. Terms are the same as in restricted recursive arithmetic. An elementary formula is T or F or a singulary predicate-variable followed by a term between ( ). Other wffs (closed and open sentences) are built up from the elementary formulas by means of connectives of the propositional calculus and both kinds of quantifiers.

(2) *Singulary second-order arithmetic with* $=$ *and* $<$ (Elgot). Differs from (1) by omission of 0 and inclusion of the two binary predicates $=$ and $<$.

(3) *First-order arithmetic with* $+$, *the binary predicate* $=$, *and a singulary predicate meaning "is a power of* 2" (Büchi).

In (1) and (2) the predicate-variables are to be understood as having *finite* sets of natural numbers as their values. And in the application to automata, in order to represent an input sequence, each free predicate-variable of a formula is made to correspond to a finite initial sequence of states of one input element by way of some suitable coding. The simplest convention, following Elgot, is just to take the value of each predicate-variable to be the finite set of instants at which the state of the corresponding input element is T, and then to consider the shortest input sequence which is thus represented. This results in a restriction to a proper subclass of all input sequences, the *admissible* input sequences, but it may be seen that the restriction is not very essential. Büchi uses (in effect) a method of coding that avoids this restriction to admissible input sequences.

In the case of (3), instead of using predicate-variables, each finite set of natural numbers is represented by a natural number, the finite sum $\Sigma 2^t$, where the summation is over all natural numbers t that belong to the set.

That (1) and (2) are equivalent, and that both have a redundant list of primitives, is as a matter of fact quite obvious. Definitions given by Elgot show that we may without loss omit the primitive symbol 0 from (1), or the primitive symbols $=$ and $<$ from (2).

The two notes by Trachtenbrot, [20] and [21], were unknown to me at the time this lecture was delivered, and were called to my attention by several members of the audience. The dates attached to the earlier note [20] place it as approximately simultaneous with [11] and the privately circulated first edition of [7], hence earlier than the abstract [3], later than [4] and the abstract of [11]—all of which it overlaps to some extent, though more in point of view and method than in specific content.

In [20] Trachtenbrot characterizes the behavior of an automaton with one binary output by means of the following formalized language:

(4) *Singulary second-order functional calculus with bounded quantifiers.* The primitive symbols are: numerical variables; singulary predicate variables; the notation ( ) for application of a propositional function to its argument; connectives of the propositional calculus; quantifiers on singulary predicate-variables; and bounded quantifiers on numerical variables. The elementary formulas consist of a singulary predicate-variable followed by a numerical variable between ( ), and other wffs are built up from the elementary formulas by means of connectives of the propositional calculus and quantifiers.

Trachtenbrot introduces four kinds of bounded quantifiers: $(a)_{a<b}$, $(\exists a)_{a<b}$, $(a)_{a\leqslant b}$, and $(\exists a)_{a\leqslant b}$, where a and b are any two distinct numerical variables. We shall say in each case that the variable a is bounded by the variable b. Evidently any one of these four kinds of bounded quantifiers would suffice, as the three others would be definable from it. The ordinary, unbounded quantifiers on numerical variables are not used.

In (4) the predicate-variables are to be understood as having arbitrary singulary propositional functions (or arbitrary sets) of natural numbers as their values. And in the application to automata, each free predicate-variable of a formula corresponds to one input element in the same fashion that was explained above for the predicates $i_1, i_2, \ldots$ used in restricted recursive arithmetic. A *t-formula* is a wff of (4) in which the only free numerical variable is the particular variable t and in which—to state it briefly—every bound numerical variable is bounded, either by t, or by a variable that is bounded by t, or by a variable that is bounded by a variable that is bounded by t, or etc. Then we have:

TRACHTENBROT'S THEOREM. The behavior of any finite automaton with one binary output $o(t)$ can be characterized by an equivalence of the form $o(t) \equiv \mathrm{S}(t)$, where $\mathrm{S}(t)$ is a t-formula whose free predicate-variables correspond to the inputs of the automaton. And conversely every equivalence of this form characterizes the behavior of some finite automaton with one binary output.

As a characterization of the behavior of finite automata with one binary output, Trachtenbrot's theorem is evidently more direct than the results of Kleene, Büchi, and Elgot. As a characterization of events representable by a finite automaton (i.e., of regular events) it is less direct. But it can be made to yield such a characterization by taking each free predicate-variable f in $\mathrm{S}(t)$ to stand for a finite initial sequence of input states of length $t+1$, namely the sequence $\mathrm{f}(0), \mathrm{f}(1), \ldots, \mathrm{f}(t)$. (This amounts to taking the formalized language in such a way that the free predicate-variables have a different range of values from the bound predicate-variables, a device which Trachtenbrot himself does not adopt explicitly in either [20] or [21].)

In order to obtain a language for practical use in specifying the behavior of automata, it may be desirable to modify Trachtenbrot's language (4) by adding the redundant primitive symbols $=$, $<$, 0, and $'$, and then to replace the class of t-formulas by the more quickly recognizable class of formulas in which the only free numerical variable is t and all bound numerical variables are bounded by t. The resulting language seems to offer less immediate facility than the regular-expression language of McNaughton and Yamada, but it may be more suitable for use in cases in which some further application of mathematical logic is to be made.

Büchi, Elgot, and Trachtenbrot are all interested, not only in applications of logic to automata theory, but also in the reverse application by which the consideration of automata is made to yield results that belong to the field of mathematical logic. And the main result announced in [21] falls under the latter head. But [21] also announces, in effect, still another characterization of regularity of events by means of expressibility in an appropriate formalized language. Namely in Büchi's characterization, weak second-order arithmetic may be replaced by another formalized language, which has the same primitive symbols and the same class of wffs as weak second-order arithmetic, but in which the values of the bound predicate-variables are taken to be arbitrary sets of natural numbers, while the values of the free predicate-variables are taken to be finite sets of natural numbers.

Finally we return to the synthesis problem, to cite some cases which belong in the present context rather than in that of restricted recursive arithmetic.

Case a. Synthesis requirement given by a regular expression, denoting an event which an automaton with one binary output is required to represent. Problem solved originally by Kleene [14], [15]. Revised treatments by Copi, Elgot, and Wright [8]; McNaughton and Yamada [17]. Elgot [10] solves a somewhat generalized form of this synthesis problem.

Case b. Synthesis requirement given by a *t*-formula that expresses the (one binary) output in terms of the inputs. Solution briefly sketched by Trachtenbrot [20].

Case c. Synthesis requirement given by a wff of the form $(\exists s_1)(\exists s_2)\ldots(\exists s_n)(t)M$, belonging to a language which is like weak second-order arithmetic except that the predicate-variables are reinterpreted as ranging over all singulary propositional functions (all sets) of natural numbers. Here $s_1, s_2, \ldots, s_n$ are predicate-variables, t is an individual variable, and the quantifier-free matrix M contains no individual variables except t. The free predicate-variables are interpreted, some as corresponding to inputs and some as corresponding to outputs, so that the synthesis requirement expressed is a condition relating the inputs and the outputs. Solved by Elgot [10].

It is clear that, as regards the form of the synthesis requirement, Case c would fit into the table that was given above in connection with restricted recursive arithmetic. But Elgot's method is different, not making use of restricted recursive arithmetic (and in fact leading in the first instance to a transition system rather than an automaton).

Case d. Synthesis requirement given by an open sentence of first-order arithmetic with $'$, $+$, and $=$, and with free singulary predicate-variables, interpreted as outputs, and as ranging over the ultimately periodic propositional functions of natural numbers. Decision problem solved, but synthesis problem unsolvable. See Elgot [10].

Case e. Synthesis requirement given by am arbitrary wff of the same formalized language that is described under Case c. The decision problem is solved by Büchi [23]. The general synthesis problem for this case is still open (as of December 1962), though unpublished results stronger than that of Case c are cited in a letter from Büchi. Evidently Case e would fit into the table that was given above, as intermediate between Cases 3–4 and Case 5.

REFERENCES

Application of propositional calculus

[A] EHRENFEST, P., Review of the Russian translation of Louis Couturat's *L'algèbre de la logique. Журнал Русскаго Физико-химическаго Общества*, section of physics, 42 (1910), part 2, 382–387.

[B] NAKASIMA, AKIRA & HANZAWA, MASAO, A series of papers published in Japan, some by the joint authors, and some by Nakasima alone. The two earliest papers are in Japanese, by Nakasima and Hanzawa, and appeared in *J. Inst. Elect. Comm. Engrs. Japan*, no. 165 (Dec. 1936) and no. 167 (Feb. 1937). I have not seen these papers, but a condensed English translation of them which appeared in *Nippon Elect. Comm. Eng.*, no. 9 (Feb. 1938), 32–39. Later papers appeared in the latter periodical, no. 10 (Apr. 1938), 178–179; no. 12 (Sept. 1938), 310–314; no. 13 (Nov. 1938), 405–412; no. 14 (Dec. 1938), 459–466; no. 24 (Apr. 1941), 203–210. These also are English translations of papers in Japanese which preceded them.

[C] SHANNON, C. E., A symbolic analysis of relay and switching circuits. *Trans. Amer. Inst. Elect. Engrs.*, 57 (1938), 713–723.

[D] SHESTAKOFF, V. I., Алгебра двухполюсных схем, построенных исключительно из двухполюсников (Алгебра А-схем). *Автоматика и Телемеханика*, no. 2 (1941), 15–24.

[E] SHESTAKOFF, V. I., A paper of the same title as [D]. *Журнал Технической Физики*, 11, 6 (1941), 532–549.

[F] SHESTAKOFF, V. I., Representation of characteristic functions of propositions by expressions realizable by relay-contact circuits. (Russian with English summary.) *Bull. Acad. Sci. URSS* (sér. math., 10 (1946), 529–554.

[G]. YANOVSKAYA, S. A., Основания математики и математическая логика. *Математика в СССР за Тридцать Лет 1917–1947*, pp. 9–50. Moscow and Leningrad, 1948.

Application of mathematical logic beyond propositional calculus

[1]. BERKELEY, E. C., *Circuit Algebra–Introduction*. New York, 1952.

[2]. BÜCHI, J. R., Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6 (1960), 66–92.

[3]. BÜCHI, J. R. & ELGOT, C. C., Decision problems of weak second order arithmetics and finite automata, Part I. Abstract in *Notices Amer. Math. Soc.*, 5, 7 (Dec. 1958), 834.

[4]. BÜCHI, J. R., ELGOT, C. C. & WRIGHT, J. B., The non-existence of certain algorithms of finite automata theory. Abstract in *Notices Amer. Math. Soc.*, 5, no. 1 (Feb. 1958), 98.

[5]. BURKS, A. W. & WRIGHT, J. B., Theory of logical nets. *Proc. I.R.E.*, 41 (1953), 1357–1365.

[6]. CHURCH, ALONZO, Review of E. C. Berkeley's "The algebra of states and events". *J. Symb. Logic*, 20 (1955), 286–287.

[7]. CHURCH, ALONZO, Application of recursive arithmetic to the problem of circuit synthesis. *Summaries of Talks Presented at the Summer Institute for Symbolic Logic, Cornell University 1957*, 2nd ed., 3–50. Princeton, 1960.

[8]. COPI, I. M., ELGOT, C. C. & WRIGHT, J. B., Realization of events by logical nets. *J. Ass. Comput. Mach.*, 5 (1958), 181–196.

[9]. ELGOT, C. C., Decision problems of weak second order arithmetics and

finite automata, Part II. Abstract in *Notices Amer. Math. Soc.*, 6, 1 (Feb. 1959), 48.

[10]. ELGOT, C. C., Decision problems of finite automata design and related arithmetics. *Trans. Amer. Math. Soc.*, 98 (1961), 21–51. Errata, ibid., 103 (1962), 558–559.

[11]. FRIEDMAN, JOYCE, Some results in Church's restricted recursive arithmetic. *J. Symb. Logic*, 22, 4 (for 1957, pub. 1958), 337–342. Reviewed by Robert McNaughton in *J. Symb. Logic*, 24, 3 (for 1959, pub. 1960), 241–242. Abstract in *J. Symb. Logic*, 21 (1956), 219.

[12]. GOTÔ, MOTINORI, Application of logical mathematics to the theory of relay networks. *Japan Sci. Rev.*, 1, 3 (1950), 35–42.

[13]. HARTREE, D. R., *Calculating Instruments and Machines*. Urbana, 1949.

[14]. KLEENE, S. C., *Representation of Events in Nerve Nets and Finite Automata*. RAND memorandum, Dec. 1951.

[15]. KLEENE, S. C., Representation of events in nerve nets and finite automata. *Automata Studies*, 3–41. Princeton, 1956.

[16]. MCCULLOCH, W. S. & PITTS, WALTER, A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5 (1943), 115–133.

[17]. MCNAUGHTON, R. & YAMADA, H., Regular expressions and state graphs for automata. *I.R.E. Trans.*, EC-9, no. 1 (1960), 39–47.

[18]. VON NEUMANN, JOHN, *First Draft of a Report on the EDVAC*. Moore School of Electrical Engineering, University of Pennsylvania, June 30, 1945.

[19]. SKOLEM, THORALF, *Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich*. Skrifter utgit av Videnskapsselskapet i Kristiania, I. matematisk-naturvidenskabelig klasse, no. 6 (1923).

[20]. TRACHTENBROT, B. A., Синтез логических сетей, операторы которых описаны средствами исчисления одноместных предикатов. *Доклады Академии Наук СССР*, 118, 4 (1958), 646–649. Reviewed by J. C. Shepherdson in *Zbl. Math.*, 84 (1960), 11–12.

[21]. TRACHTENBROT, B. A., Некоторые построения в логике одноместных предикатов. *Доклады Академии Наук СССР*, 138, 2 (1961), pp. 320–321. English translation in *Soviet Math.*, 2, 3 (1961), 623–625.

[22]. WANG, HAO, Circuit synthesis by solving sequential Boolean equations. *Z. Math. Logik Grundlagen Math.*, 5 (1959), 291–322. Reviewed by C. C. Elgot in *J. Symb. Logic*, 25, 4 (for 1960, pub. 1962), 373–375.

(Added in proof):

[23]. BÜCHI, J. R., On a decision method in restricted second-order arithmetic. *Logic, Methodology, and Philosophy of Science, Proc. of the 1960 Internat. Congress*, Stanford 1962, 1–11.

МАРКОВСКИЕ ПРОЦЕССЫ И ЗАДАЧИ АНАЛИЗА

Е. Б. ДЫНКИН

Связи между марковскими процессами и дифференциальными уравнениями известны уже давно. Еще в начале тридцатых годов они были объектом глубоких исследований Колмогорова, Петровского, Хинчина. За последние годы были обнаружены новые замечательные связи между марковскими процессами и классическим анализом. Эти связи плодотворны не только для теории вероятностей, но и для анализа. Современная теория марковских процессов становится все более и более не только разделом теории вероятностей, но и важной областью анализа, оказывающей влияние на развитие таких классических областей как теория эллиптических и параболических дифференциальных уравнений, теория потенциала и т. п.

О широком интересе к этой области свидетельствует место, которое уделено ей в программе нашего съезда, так же, как и в программе предыдущего международного съезда в Эдинбурге. Сыгравшие выдающуюся роль в новейшем развитии теории марковских процессов исследования Феллера были освещены автором в часовом докладе на Эдинбургском конгрессе. Важные направления в теории марковских процессов связаны с именами Ито и Ханта, которые выступают с обзорными докладами на нашем съезде.

Это облегчает мою задачу. Я не буду пытаться охватить все аспекты современного развития теории марковских процессов (задача, неразрешимая в рамках часового доклада), а сосредоточу внимание на нескольких направлениях, связанных с моей собственной работой и работой группы молодых советских математиков, находящихся со мной в постоянном научном контакте.

Речь будет идти о понятии характеристического оператора марковского процесса, позволяющем дать вероятностную формулировку и вероятностное решение ряда задач теории дифференциальных уравнений. Значительное место будет уделено понятию аддитивного функционала от марковского процесса и различным преобразованиям марковских процессов, связанным с аддитивными функционалами. Будет показано, что с помощью таких преобразований можно получить некоторое представление о строении наиболее общего строго марковского процесса с непрерывными траекториями на топологическом многообразии E. Мы введем определенные классы функций, которые естественно назвать гармоническими и супергармоническими функциями, связанными с марковским процессом, и остановимся на некоторых задачах, возникающих в связи с изучением множества всех неотрицательных гармонических функций.

§ 1. Современное определение марковского процесса

1.1. В 1828 году английский ботаник Броун заметил, что мелкие частицы, погруженные в жидкость, хаотически движутся, беспрестанно меняя направление движения. В первых работах, посвященных математическому описанию этого явления, изучалась так называемая переходная функция $P(t, x, \Gamma)$, выражающая вероятность того, что частица, вышедшая из точки x, через время t попадает в множество Γ.

Более развитая теория имеет своим объектом траекторию движения x_t. Случайный характер движения выражается математически предположением, что $x_t = x_t(\omega)$, где ω принадлежит некоторому множеству Ω, на котором задан набор вероятностных мер P_x (это множество называется пространством элементарных событий). Множества A, для которых определены значения $P_x(A)$, называются событиями, и значение $P_x(A)$ истолковывается как вероятность события A в предположении, что движение начинается из точки x. Переходная функция определяется формулой $P(t, x, \Gamma) = P_x\{x_t \in \Gamma\}$.

1.2. Мы приходим таким образом к современному определению марковского процесса как пары (x_t, P_x), где $x_t = x_t(\omega)$ — функция, заданная при $t \geqslant 0$ и $\omega \in \Omega$, и P_x — набор вероятностных мер в пространстве Ω. *Фазовое пространство*, в котором принимает значения функция x_t, является в случае броуновского движения некоторой областью трехмерного евклидова пространства. Вообще же это произвольное множество E, в котором выделена некоторая система «измеримых подмножеств». Основное условие, связывающее функцию x_t и меры P_x, — это марковский принцип независимости будущего от прошлого при известном настоящем. Точнее, при известном значении x_t прогноз дальнейшего движения частицы не зависит от характера движения до момента t (1).

Разобьем всю траекторию процесса на две части: до момента τ первого достижения некоторого множества Γ и после этого момента. Предположим, что нам известно x_τ. Существенно ли знание движения до момента τ для прогноза движения после момента τ? Физическая интуиция требует отрицательного ответа. Однако такой ответ вовсе не вытекает из определения марковского процесса, в котором участвует фиксированный момент t, а не случайный момент τ. Марковские процессы, для которых условие независимости будущего от прошлого при известном настоящем выполняется не только для постоянного, но и для определенного класса случайных моментов τ, называются *строго марковскими процессами* (2).

---

(1) Строгая формулировка опирается на понятие условной вероятности и требует, чтобы для любого измеримого множества Γ и любых $t, h \geqslant 0$ $P_x\{x_{t+h} \in \Gamma / x_s, s \leqslant t\} = P(h, x_t, \Gamma)$ при почти всех ω.

(2) Изучение строго марковских процессов как самостоятельного класса марковских процессов было начато в 1955–1956 годах в работах Дынкина [5], [6], [8], Дынкина и Юшкевича [18] и Рэя [38]. В работе [18] впервые дано общее определение строго марковских процессов, строятся примеры марковских процессов, не являющихся строго марковскими, и выводятся условия, достаточные для того, чтобы марковский процесс являлся строго марковским.

1. 3. До сих пор мы считали, что x_t определено для всех $t \geqslant 0$. Однако, многие естественные конструкции приводят к процессам, для которых $x_t(\omega)$ определено лишь в некотором интервале $(0, \zeta(\omega))$. Современная теория марковских процессов охватывает и такие обрывающиеся процессы.

Определение марковского процесса не требует, чтобы в фазовом пространстве была задана какая-нибудь топология. Однако в развитой теории широко используются условия топологического характера и ту или иную топологию в фазовом пространстве приходится вводить. Среди различных топологий особое место занимает так называемая естественная топология, в которой открытые множества выделяются следующим условием: траектория, начинающаяся в произвольной точке такого множества, с вероятностью единица не выходит из этого множества в течении положительного времени (см. Дуб [2], Хант [28], Дынкин [10]). Для широкого класса процессов доказано, что измеримая функция $f(x)$ является непрерывной в естественной топологии тогда и только тогда, когда она непрерывна справа вдоль почти всех траекторий процесса (1).

§ 2. Операторы сдвига функций. Характеристические и инфинитезимальные операторы

2. 1. Общая схема, связывающая марковские процессы с анализом, основана на понятии сдвига функции, заданной в фазовом пространстве E. Рассмотрим какую-нибудь неотрицательную измеримую функцию $\tau(\omega)$. Пусть $f(x)$ — измеримая функция в пространстве E. Тогда $f(x_\tau)$ является функцией в пространстве Ω. Интеграл этой функции по мере P_x (если он имеет смысл) и есть значение сдвинутой функции в точке x. В виде формулы это записывается так

$$T_\tau f(x) = M_x f(x_\tau) \text{ (2)}.$$

Рассмотрим внимательнее случай, когда $\tau = t$ не зависит от ω. Соответствующий оператор сдвига выражается через переходную функцию по формуле

$$T_t f(x) = \int_E P(t, x, dy) f(y).$$

Из марковского принципа вытекает, что $T_s T_t = T_{s+t} (s, t \geqslant 0)$, т. е. операторы T_t образуют полугруппу. Оператор

$$A f(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t}$$

Дальнейшие указания относительно литературы по строго марковским процессам можно найти в [11].

(1) Этот результат принадлежит Гирсанову [23], опиравшемуся на некоторые идеи Дуба [2].

(2) Через $M_x \xi$ обозначается интеграл функции ξ по мере P_x, распространенный на всю область определения функции $\xi(\omega)$.

называется *инфинитезимальным оператором* марковского процесса. При очень широких предположениях инфинитезимальный оператор определяет однозначно переходную функцию процесса (см. [7], [17]).

Другой важной характеристикой процесса является характеристический оператор, который определяется формулой

$$\mathfrak{A} f(x) = \lim_{U \downarrow x} \frac{T_{\tau(U)} f(x) - f(x)}{M_x \tau(U)}.$$

Здесь U — окрестность точки x, $\tau(U)$ — момент первого выхода из U, предельный переход производится, когда U стягивается к x. (Предполагается, что в фазовом пространстве задана некоторая топология.)

Между характеристическим и инфинитезимальным оператором марковского процесса существует тесная связь. Именно, если в некоторой топологии траектории процесса непрерывны справа, а операторы сдвига T_t оставляют инвариантным множество всех ограниченных непрерывных функций, то характеристический оператор $\mathfrak{A}$ является расширением инфинитезимального оператора A (1).

Из формулы, определяющей оператор $\mathfrak{A}$, видно, что если $\mathfrak{A} f(x)$ определено и функция f достигает в точке x неотрицательного максимума, то $\mathfrak{A} f(x) \leqslant 0$. Это свойство часто называют *принципом максимума*.

Начиная с этого места мы сосредоточим все внимание на процессах, траектории которых непрерывны. Для таких процессов в момент первого выхода из множества U частица попадает на границу U. Поэтому значение $\mathfrak{A} f(x)$ определяется по значениям функции f в сколь угодно малой окрестности точки x. Оператор $\mathfrak{A}$ является *локальным* линейным оператором.

§ 3. Диффузионные процессы. Вероятностное решение дифференциальных уравнений

3. 1. Строго марковский процесс с непрерывными траекториями условимся называть *диффузионным процессом*, если $\mathfrak{A} f(x)$ определено для каждой функции f, дважды непрерывно дифференцируемой в окрестности точки x. (Предполагается, что в фазовом пространстве введена структура гладкого многообразия.) Доказывается, что для диффузионного процесса характеристический оператор совпадает на дважды непрерывно дифференцируемых функциях с некоторым дифференциальным оператором второго порядка

$$L f(x) = \sum a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum b_i(x) \frac{\partial f(x)}{\partial x_i} - c(x) f(x).$$

Из принципа максимума вытекает, что $c(x) \geqslant 0$ и выполняется ослабленное условие эллиптичности: $\Sigma a_{ij}(x) \lambda_i \lambda_j \geqslant 0$ для любых вещественных $\lambda_1, \lambda_2, \ldots$. Оператор L называется *производящим дифференциальным*

(1) Понятие характеристического оператора и связь между характеристическими и инфинитезимальными операторами были впервые изучены в работах [5], [8].

оператором диффузионного процесса. (Он является сужением характеристического оператора.) Величины a_{ij}, b_i и c называются, соответственно, коэффициентами диффузии, коэффициентами сноса и коэффициентом обрыва (если процесс не обрывается, то $c(x)=0$).

3. 2. Важно научиться строить диффузионные процессы, отправляясь от дифференциальных операторов L. Диффузионный процесс в евклидовом пространстве, отвечающий оператору Лапласа Δ, был построен еще в 1923 году Винером. Мы будем называть его *винеровским процессом*. Конструкцию Винера можно распространить на любой эллиптический оператор L, для которого удается построить фундаментальное решение параболического дифференциального уравнения

$$\frac{\partial u_t}{\partial t}=Lu_t.$$

В случае евклидова пространства для этого достаточно, чтобы коэффициенты оператора L были ограничены и удовлетворяли условию Гельдера и чтобы существовала положительная нижняя граница для собственных значений матрицы коэффициентов при старших производных ([1]).

Более простой, вероятностный метод построения необрывающихся диффузионных процессов был предложен Ито [30], [31].

Пусть (x_t, P_x) — одномерный винеровский процесс и $(\tilde{x}_t, \tilde{P}_x)$ — диффузионный процесс с производящим дифференциальным оператором

$$L=\frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}+m(x)\frac{d}{dx}.$$

Ито показал, что оба процесса можно реализовать на одном и том же пространстве элементарных событий Ω так, чтобы $\tilde{P}_x=P_x$ и траектории $\tilde{x}_t$ и x_t были связаны соотношением

$$\tilde{x}_t=\tilde{x}_0+\int_0^t m(x_u)\,du+\int_0^t \sigma(x_u)\,dx_u \tag{1}$$

(справа стоит так называемый стохастический интеграл, общее определение которого было дано Ито). Формулу (1) можно рассматривать как интегральное уравнение, позволяющее выразить $\tilde{x}_t$ через x_t. Аналогичное уравнение пишется и для многомерных процессов.

Уравнение (1) можно записать в дифференциальной форме

$$d\tilde{x}_t=m(x_t)\,dt+\sigma(x_u)\,dx_u.$$

Последнее соотношение представляет собой стохастический аналог обыкновенного дифференциального уравнения и обладает многими свойствами таких уравнений: его можно решать методом последовательных приближений или методом ломанных Эйлера, его исследование нисколько не

([1]) См. Погожельский [36], [37].

усложняется с увеличением числа измерений пространства. Еще одним замечательным преимуществом метода Ито является его нечувствительность к вырождению исходного дифференциального оператора.

3. 3. Рассмотрим несколько типичных задач теории дифференциальных уравнений, связанных с оператором L:

А. (Задача Дирихле.) $Lf(x)=0$ для $x\in G, f(x)=\varphi(x)$ для $x\in G'$.

Б. $Lf(x)=-g(x)$ для $x\in G, f(x)=0$ для $x\in G'$.

В. $\dfrac{\partial u_t(x)}{\partial t}=Lu_t(x)$ для $t>0, x\in G; u_0(x)=v(x)$ для $x\in G; u_t(x)=0$ для $x\in G'$.

Здесь G означает некоторую область в евклидовом пространстве, G' — ее границу, φ, g, v — известные функции, f и u_t — неизвестные функции, которые надо найти.

Заменив в формулировках задач А, Б, В оператор L на характеристический оператор $\mathfrak{A}$, мы получим новые задачи А′, Б′, В′. Поскольку $\mathfrak{A}\supseteq L$, то всякое решение какой-нибудь из задач А, Б, В является в то же время решением соответствующей задачи А′, Б′, В′. Естественно назвать решения задач А′, Б′, В′ *обобщенными решениями* задач А, Б, В.

Из принципа максимума для оператора $\mathfrak{A}$ вытекает, что при широких предположениях задачи А′, Б′, В′ имеют не более одного решения. Если это решение дважды непрерывно дифференцируемо, то она является решением соответствующей задачи А, Б, В. Таким образом, в этом случае обобщенное решение задачи А, Б, В является «классическим решением».

Замечательно, что решение задач А′, Б′, В′ всегда выражается явными вероятностными формулами. Именно, обозначая через β момент первого выхода области G, имеем:

для А′: $f(x)=M_x\varphi(x_\beta)$

для Б′: $f(x)=M_x\int_0^\beta g(x_t)\,dt$

для В′: $u_t(x)=M_x v(x_t)\chi_{\beta<t}$ $\left(\chi_A(\omega)=\begin{cases}1 \text{ при } \omega\in A\\ 0 \text{ при } \omega\notin A\end{cases}\right)$

Эти формулы позволяют исследовать решение с качественной стороны, в частности, исследовать его зависимость от начальных и граничных данных и правых частей (см., например, Хасьминский [24]–[27], Фрейдлин [21], [22]). Кроме того они могут служить основой для вычислений по методу Монте-Карло.

Аналогичные формулы можно дать и для более сложных уравнений. Например, обобщенное решение дифференциального уравнения

$$Lf-Vf=-g$$

при нулевых граничных условиях выражается формулой

$$f(x)=M_x\int_0^\beta \exp\left(-\int_0^t V(x_u)\,du\right)g(x_t)\,dt.$$

3. 4. В задаче В мы рассматривали простейшее, нулевое граничное условие для параболического дифференциального уравнения $\partial u_t/\partial t=Lu_t$. Известно, что для этого уравнения можно ставить и другие граничные условия, например, можно потребовать, чтобы обращалась в нуль нормальная производная $\partial u_t/\partial n$. Различным граничным условиям соответствуют различные типы поведения движущейся частицы на границе области. Например, нулевым граничным условиям соответствует исчесновение в момент первого достижения границы; условию $\partial u_t/\partial n=0$ соответствует отражение по направлению нормали.

Задача определения всех возможных граничных условий для диффузионных процессов была изучена в одномерном случае Феллером [19], [20] и в многомерном случае Вентцелем [47], [48].

Задачу можно уточнить следующим образом. Пусть G — область с гладкой границей G' и L — дифференциальный оператор в G, коэффициенты которого продолжаются непрерывно на $G\cup G'$. Требуется описать все диффузионные процессы в замкнутой области $G\cup G'$, для которых производящий дифференциальный оператор во всех внутренних точках совпадает с L. Пока еще не было определено, что значит, что процесс является диффузионным в граничной точке x. Так же, как и для внутренней точки это понятие определяется через характеристический оператор. Требуется, чтобы значение $\mathfrak{A}f(x)$ было определено либо (а) для всех гладких функций, либо (б) для всех гладких функций, подчиненных одному линейному условию (под гладкими функциями понимаются здесь функции, дважды непрерывно дифференцируемые в некоторой окрестности точки x).

Вентцель показал (1), что в случае (а) на гладких функциях

$$\mathfrak{A}f(x)=-qf+\gamma\frac{\partial f}{\partial n}+L'f,$$

где L' — дифференциальный оператор второго порядка на G'(2); в случае (б) линейное условие, задающее область определения имеет вид

$$0=-qf+\gamma\frac{\partial f}{\partial n}+L'f.$$

Если функция $\mathfrak{A}f$ непрерывна, то в первой из написанных формул можно заменить $\mathfrak{A}$ на L, и обе формулы записываются в виде

$$-\sigma Lf-qf+\gamma\frac{\partial f}{\partial n}+L'f=0.$$

---

(1) Результаты Вентцеля излагаются в переработанном виде.

(2) Мы считаем, что траектории процесса непрерывны. Феллер и Вентцель допускали возможность разрывов на границе. Поэтому у них появляются дополнительные интегральные члены.

Итак, для того, чтобы задать граничное условие, надо определить на границе функции σ, q, γ и дифференциальный оператор L'. При $L' \equiv 0$ получаем граничные условия, хорошо известные в теории дифференциальных уравнений. Теоремы существования и единственности для общего случая разбирались Вентцелем и Уено [44]. Однако, полученные в этом направлении результаты еще никак нельзя считать окончательными.

С наглядной точки зрения сформулированный выше результат означает, что единственно возможными типами поведения диффундирующей частицы на границе области являются остановка, исчезновение, отражение, диффузия по границе и их различные комбинации. Слово «комбинация» означает просто линейную комбинацию соответствующих граничных условий, но вероятностный смысл такого комбинирования совсем не прост.

Каждому типу граничных условий соответствует определенный процесс, происходящий на границе. Он определен на случайном множестве моментов, в которые частица находится на границе. Изучение граничных процессов можно поэтому рассматривать как одну из задач, еще не построенной общей теории марковских процессов со случайной областью определения. Построение такой теории представляется весьма интересной проблемой. (Интересные соображения по этому поводу были высказаны в одном из докладов Колмогорова на секции теории вероятностей и математической статистики Московского математического общества.)

§ 4. Аддитивные функционалы

4. 1. За последние годы одним из важнейших орудий исследования и применения марковских процессов становятся аддитивные функционалы от этих процессов.

Аддитивный функционал от марковского процесса (x_t, P_x) представляет собой функцию, сопоставляющую каждому интервалу времени $[s,t]$ случайную величину φ_t^s зависящую лишь от течения процесса за время $[s,t]$ (1). При этом требуется, чтобы $\varphi_t^s + \varphi_u^t = \varphi_u^s$ для любых $s \leqslant t \leqslant u$. Мы будем рассматривать только аддитивные функционалы, удовлетворяющие следующему условию однородности: при замене траектории x_t на x_{t+h} значение φ_t^s заменяется на φ_{t+h}^{s+h} (2).

Примером аддитивного функционала является

$$\varphi_t^s = \int_s^t V(x_u)\,du,$$

где $V(x)$ — какая-нибудь измеримая функция в фазовом пространстве. Если функция V неотрицательна, то функционал φ_t^s принимает только неотрицательные значения. Такие функционалы представляют особый интерес для применений. В 1958 г. Волконский [45] (см. также [46]) описал все непрерывные неотрицательные аддитивные функционалы от

(1) Точнее, функция $\varphi_t^s(\omega)$ измерима относительно σ-алгебры, порожденной функциями $x_u(\omega)$ при $u \in [s,t]$.

(2) Точнее, если для некоторых ω_1 и ω_2 из Ω и $h > 0$ $x_t(\omega_1) = x_{t+h}(\omega_2)$ при всех t, то $\varphi_t^s(\omega_1) = \varphi_{t+h}^{s+h}(\omega_2)$ при всех $s \leqslant t$.

одномерного винеровского процесса. Оказалось, что эти функционалы находятся во взаимно однозначном соответствии с мерами на прямой, принимающими конечные значения на каждом конечном отрезке. Это соответствие непрерывно, если рассматривать для мер слабую сходимость, а для функционалов сходимость по вероятности. Функционал, соответствующий мере μ, записывается в виде

$$\varphi_t^s = \int_s^t \frac{d\mu}{dx}(x_u)\,du. \tag{2}$$

При этом $d\mu/dx$ означает плотность меры μ относительно меры Лебега. Если такой плотности не существует, то запись носит символический характер и расшифровывается посредством естественного предельного перехода.

Для многомерного винеровского процесса задача была решена в 1960 г. Вентцелем [49] и Маккином и Танака [33]. Так же, как и в одномерном случае, имеет место однозначное соответствие между неотрицательными аддитивными функционалами и некоторым классом мер в фазовом пространстве. Однако, этот класс мер описывается сложнее. (Он содержит, в частности, все меры, обладающие ограниченной плотностью относительно меры Лебега, а также все меры с ограниченным ньютоновым потенциалом.) Непрерывность соответствия между мерами и функционалами больше не имеет места, однако интегралу, стоящему в правой части формулы (2) все же можно придать смысл с помощью более осторожного предельного перехода.

Не исключено, что теорему о непрерывности соответствия между мерами и функционалами можно спасти, если при определении слабой сходимости мер рассматривать не обычную топологию евклидова пространства, а естественную топологию, связанную с винеровским процессом. Было бы интересно выяснить, верно ли это предположение.

4.2. Примером непрерывного аддитивного функционала, принимающего значения разных знаков, может служить

$$\varphi_t^s = f(x_t) - f(x_s), \tag{3}$$

где f — произвольная непрерывная функция. Дынкин [13] показал, что непрерывные аддитивные функционалы от винеровского процесса можно получить с помощью стохастических интегралов по формуле

$$\varphi_t^s = \int_s^t b(x_u)dx_u. \tag{4}$$

Опираясь на некоторые идеи Скорохода [42] Вентцель [50], [51] доказал, что всякий непрерывный аддитивный функционал от одномерного винеровского процесса представляется в виде суммы функционалов вида (3) и (4). Аналогичный результат получен им и для многомерного случая, но его формулировка несколько сложнее.

§ 5. Супергармонические и гармонические функции

5. 1. Супергармонические и гармонические функции играют важную роль в классическом анализе.

Одно из возможных определений супергармонической функции требует, чтобы функция была непрерывна сверху и чтобы среднее ее значение по любой сфере не превосходило ее значения в центре сферы.

Пусть τ — момент первого выхода траектории винеровского процесса из сферы S с центром в точке x. Очевидно, точка x_τ равномерно распределена по сфере S, и поэтому среднее значение функции f по S равно $T_\tau f(x) = M_x f(x_\tau)$. Таким образом, основное неравенство для супергармонических функций можно записать в виде $T_\tau f(x) \leqslant f(x)$. Доказывается, что это неравенство сохраняет силу, если сфера с центром в x заменяется произвольной компактной окрестностью точки x. Доказывается также, что супергармонические функции непрерывны в естественной топологии и ограничены снизу на каждом компакте. Более того, оказывается, что перечисленные требования (непрерывность в естественной топологии, ограниченность снизу на каждом компакте и неравенство $T_\tau f(x) \leqslant f(x)$ для каждой компактной окрестности точки x) равносильны супергармоничности функции f. Мы приходим к вероятностному определению супергармонических функций.

Это определение сохраняет смысл при замене винеровского процесса на любой марковский процесс в локально компактном фазовом пространстве. Таким образом, для каждого марковского процесса можно ввести класс связанных с ним супергармонических функций ([1]). Для процессов с непрерывными траекториями это определение естественно локализуется, так что можно говорить о функциях, супергармонических в некоторой области G.

5.2. В классическом анализе заметное место занимает задача о представлении супергармонических функций в виде потенциалов. Ньютонов потенциал меры μ в трехмерном евклидовом пространстве E представляет собой функцию $f(x)$, определяемую формулой

$$f(x) = \int_E \frac{\mu(dy)}{|y-x|}$$

($|y-x|$ означает длину вектора $y-x$). Эта функция неотрицательна и супергармонична.

Нетрудно доказать, что если φ_t^s — аддитивный функционал от винеровского процесса, отвечающий мере μ, то потенциал дается формулой

$$f(x) = M_x \varphi_\infty^0. \tag{5}$$

Естественно поставить общую задачу о представлении неотрицательных супергармонических функций, связанных с произвольным марковским процессом, в виде (5). Для широкого класса процессов эта задача решена

([1]) Приведенное определение принадлежит докладчику [10], [17]. Как показано в [10] класс всех неотрицательных супергамонических функций совпадает с классом эксцессивных функций, введенных Хантом [28].

в работах Волконского [46], Вентцеля [49], Мейера [35], Шура [39], [40]. Доказано, что представление (5) возможно для всех ограниченных неотрицательных супергармонических функций. Получены критерии представимости для неограниченных функций.

5. 3. Функция f называется *гармонической*, если функции f и $-f$ являются супергармоническими. Гармонические функции являются решениями уравнения $\mathfrak{A}f(x)=0$, где $\mathfrak{A}$ — характеристический оператор процесса. (В случае винеровского процесса они являются решениями уравнения Лапласа $\Delta f(x)=0$.)

Пусть (x_t, P_x) — необрывающийся непрерывный строго марковский процесс на n-мерном топологическом многообразии E. Условимся говорить, что функции $f_1, \ldots, f_n$ представляют собой *гармонические координаты* вблизи x, если эти функции являются гармоническими в некоторой окрестности U точки x и определяют топологическое отображение U на некоторую область евклидова пространства. Если кроме того всякая функция, гармоническая в U, является функцией класса C^k (1) от $f_1, \ldots, f_n$, то мы будем говорить, что функции $f_1, \ldots, f_n$ суть гармонические координаты класса C^k.

Нетрудно видеть, что если вблизи каждой точки x из E можно ввести гармонические координаты класса C^k, то набор всех таких координат определяет в E структуру гладкого многообразия класса C^k. Мы назовем ее *естественной гладкой структурой*.

Условимся говорить, что марковский процесс *регулярен в точке* x, если вблизи x можно ввести гармонические координаты. В одномерном случае это условие равносильно требованию, чтобы существовал интервал $(a,b)\supset x$ со следующим свойством: траектория, начинающаяся из любой точки $y\in(a,b)$ попадает с положительной вероятностью в каждую из точек a,b в момент первого выхода из (a,b). Интуитивно ясно, что в многомерном случае условие регулярности имеет аналогичный смысл. Было бы интересно найти соответствующую точную формулировку.

Очень интересно было бы выяснить, насколько широк класс процессов, допускающих естественную гладкую структуру класса C^k. Этим свойством во всяком случае обладают (при $k=\omega$) все регулярные одномерные процессы, а также все диффузионные процессы, для которых производящий дифференциальный оператор является эллиптическим оператором с аналитическими коэффициентами.

Понятие регулярности процесса и понятие естественной гладкой структуры были введены нами для необрывающихся процессов. Если процесс может обрываться, эти определения необходимо видоизменить. Для одномерного случая корректное определение естественной гладкой структуры дано в [17]. В многомерном случае такое определение еще должно быть найдено.

---

(1) Через C^k ($k=0, 1, 2, \ldots, \infty$) обозначается класс всех k раз непрерывно дифференцируемых функций. Через C^ω обозначается класс всех вещественных аналитических функций.

§ 6. Преобразования марковских процессов, связанные с аддитивными функционалами

6. 1. Знание аддитивных функционалов от марковского процесса позволяет строить по этому процессу ряд новых марковских процессов.

Мы рассмотрим несколько операций над марковскими процессами. Одна из них — построение подпроцесса (1). Эта операция состоит в том, что траектория процесса x_t обрывается в некоторый момент $\tilde{\zeta}$. При заданной траектории момент $\tilde{\zeta}$ остается случайным, и для того, чтобы определить подпроцесс, надо задать соответствующее условное распределение вероятностей. Оказывается, что все эти условные распределения связаны с аддитивными функционалами от исходного процесса. Именно, если φ_t^s — аддитивный функционал, неотрицательный и непрерывный относительно t, то можно построить подпроцесс, для которого условная вероятность события $\tilde{\zeta}>t$ равна $e^{-\varphi_t^s}$.

Рассмотрим простейший случай, когда функционал φ_t^s задается формулой

$$\varphi_t^s=\int_s^t V(x_u)\,du \tag{6}$$

($V(x)$ — непрерывная неотрицательная функция.) Легко подсчитать, что в этом случае условная вероятность события $t\leqslant\tilde{\zeta}<t+dt$ равна $V(x_t)dt+o(dt)$ и, следовательно, траектория, находящаяся в некоторый момент t в точке x, обрывается за время dt с вероятностью $V(x)dt+o(dt)$ независимо от характера движения до момента t. Если оборвать таким образом диффузионный процесс с производящим дифференциальным оператором L, то получится диффузионный процесс с производящим оператором

$$\tilde{L}f=Lf-Vf\,(2).$$

Функцию $V(x)$ естественно назвать *плотностью вероятности обрыва*. Для широкого класса процессов эта плотность может быть определена для каждого подпроцесса, только, вообще говоря, она является обобщенной функцией.

Поясним это утверждение для случая винеровского процесса. Для такого процесса каждый неотрицательный аддитивный функционал характеризуется некоторой мерой μ. Следовательно, каждый подпроцесс винеровского процесса также характеризуется некоторой мерой μ. Производная $d\mu/dx$ меры μ относительно меры Лебега является обобщенной функцией, и эта функция представляет собой естественное обобщение введенной выше плотности вероятности обрыва. Меру μ мы назовем *убивающей мерой*: если для некоторой области G $\mu(G)=0$, то траектория процесса с вероятностью 1 не обрывается в области G; на-

(1) См. [11], [17].

(2) Этот результат легко следует из общей теоремы устанавливающей аналогичную связь инфинитезимальными операторами процесса и подпроцесса (см. [4]).

против, если значение $\mu(G)$ велико, то частица имеет мало шансов пройти область G, не погибнув.

6.2. Рассмотрим далее преобразование марковского процесса (x_t, P_x), при котором траектории не меняются, а меры преобразуются по формуле

$$\tilde{P}_x(d\omega) = \xi(\omega) P_x(d\omega).$$

Чтобы пара $(x_t, \tilde{P}_x)$ задавала марковский процесс, достаточно положить

$$\xi = e^{-\varphi_\infty^s},$$

где φ_t^s $(0 \leqslant s \leqslant t \leqslant \infty)$ — аддитивный функционал, удовлетворяющий условию

$$M_x e^{-\varphi_\infty^s} = 1.$$

Для винеровского процесса функционалы φ, удовлетворяющие последнему условию, можно строить по формуле

$$\varphi_t^s = \int_s^t b(x_u)\, dx_u + \frac{1}{2}\int_s^t b(x_u)^2\, du.$$

(первый член представляет собой стохастический интеграл, функция $b(x)$ принимает векторные значения и $b(x)^2$ означает ее скалярный квадрат). При преобразовании мер, отвечающем этому функционалу, из винеровского процесса получается диффузионный процесс с производящим дифференциальным оператором

$$\tilde{L}f = \Delta f - \sum b_i(x) \frac{\partial f}{\partial x_i}$$

($b_i(x)$ — декартовы координаты вектора $b(x)$).

6.3. В каждом из рассмотренных нами случаев построение нового марковского процесса происходит с помощью некоторого аддитивного функционала φ_t^s. Простая выкладка показывает, что переходная функция преобразованного процесса задается каждый раз одной и той же формулой

$$\tilde{P}(t, x, \Gamma) = \int_{x_t \in \Gamma} e^{-\varphi_t^s} P_x(d\omega)\,.$$

В настоящее время известен обширный класс преобразований, обладающих тем же свойством (см. [14], [17]). При каждом из этих преобразований обрывается траектория процесса и в то же время определенным образом преобразуются меры P_x. К числу наиболее важных примеров таких преобразований относится преобразование, связанное с аддитивным функционалом

$$\varphi_t^s = -\log \frac{f(x_t)}{f(x_0)},$$

где f — положительная супергармоническая функция (1). При этом преобразовании переходная функция преобразуется по формуле

$$\tilde{P}(t, x, \Gamma) = \frac{1}{f(x)} \int_{\Gamma} P(t, x, dy) f(y).$$

6. 4. Весьма важное место занимает в теории марковских процессов *случайная замена времени*, впервые систематически изученная Волконским [45], [46] (см. также Ито и Маккин [32]). Эта операция состоит в том, что вдоль каждой траектории $x_t(\omega)$ вводится новое время $\tau_t(\omega)$, другими словами, вместо процесса (x_t, P_x) рассматривается пара (x_{τ_t}, P_x). Функция $\tau_t(\omega)$ должна быть, конечно, непрерывной и возрастающей функцией от t. Возникает вопрос: как выбрать среди таких функций все функции, для которых пара (x_{τ_t}, P_x) определяет марковский процесс. Оказывается, для этого достаточно рассмотреть всевозможные аддитивные функционалы $\varphi_t^s(\omega)$ от процесса (x_t, P_x), непрерывные относительно t и положительные при любых $t > s \geqslant 0$, и построить для каждого такого функционала функцию $\tau_t(\omega)$, обратную для функции $\varphi_t^0(\omega)$. Простейшим примером аддитивного функционала, удовлетворящего всем необходимым условиям, является функционал (6) при положительной функции $V(x)$. Значение τ_t находится из соотношения

$$\int_0^{\tau_t} V(x_u) du = t.$$

Ясно, что $d\tau_t/dt = V(x_{\tau_t})^{-1}$ и, следовательно, замена времени сводится к тому, что в каждой точке x движение ускоряется (или замедляется) с коэффициентом $V(x)$.

Вообще при случайной замене времени «карта траекторий» процесса не меняется. Поэтому не меняются вероятности, характеризующие положение частицы в момент первого выхода из любого открытого множества G (мы называем их для краткости *вероятностями выхода*). Естественно спросить, вытекает ли из совпадения вероятностей выхода для двух марковских процессов, что каждый из них может быть получен из другого посредством случайной замены времени. В случае, когда один из процессов — винеровский, положительный ответ на этот вопрос дается в работе Маккина и Танака [33]. В работе Блюменталя, Гетура и Маккина [1] этот результат распространяется на широкий класс процессов.

Пусть $\mathfrak{A}$ — характеристический оператор некоторого процесса, и $\tilde{\mathfrak{A}}$ — характеристический оператор процесса, полученного посредством случайной замены времени. В формуле, определяющей $\mathfrak{A}f(x)$ числитель выражается через f и вероятности выхода, а знаменатель не зависит от f. Поэтому для всех функций f, входящих одновременно в области определения $\mathfrak{A}$ и $\tilde{\mathfrak{A}}$, отношение $\mathfrak{A}f(x)/\tilde{\mathfrak{A}}f(x)$ принимает одно и то же значение. Обозначая его через $V(x)$, имеем формулу

$$\tilde{\mathfrak{A}} f(x) = V(x)^{-1} \mathfrak{A} f(x).$$

(1) Это преобразование строится при дополнительном предположении, что f представимо в виде (5).

Опираясь на эту формулу, нетрудно доказать, что диффузионный процесс с производящим дифференциальным оператором $\tilde{L}$ можно получить посредством случайной замены времени из диффузионного процесса с производящими дифференциальным оператором L тогда и только тогда, когда $\tilde{L}=V^{-1}L$, где $V(x)$ — некоторая положительная функция. При этом замена времени может быть осуществлена с помощью аддитивного функционала (6).

§ 7. Обобщенное броуновское движение

7. 1. Условимся называть *обобщенным броуновским движением* всякий марковский процесс, который можно получить из винеровского процесса посредством операций образования подпроцесса и случайной замены времени. Любое число таких операций можно заменить одной операцией образования подпроцесса и одной операцией случайной замены времени. Как уже говорилось, каждая из этих двух операций определяется аддитивным функционалом. Эти функционалы находятся во взаимно однозначном соответствии с определенными классами мер. Таким образом, всякое обобщенное броуновское движение характеризуется двумя мерами: первая — *«убивающая мера»* μ задает, каким образом винеровский процесс обрывается; вторая — *«мера скорости»* ν определяет случайную замену времени.

Нетрудно видеть, что вероятности выхода обобщенного броуновского движения мажорируются вероятностями выхода винеровского процесса. Как показал недавно Шур, для весьма широкого класса процессов это требование не только необходимо, но и достаточно для того, чтобы процесс являлся обобщенным броуновским движением.

7. 2. Рассмотрим броуновское движение с убивающей мерой μ и мерой скорости ν. Если существуют непрерывные производные $d\mu/d\nu$ и $d\nu/dx>0$, то этот процесс является диффузионным и его производящий дифференциальный оператор равен

$$Lf=\frac{1}{\frac{d\nu}{dx}}\Delta f-\frac{d\mu}{d\nu}f. \tag{7}$$

В общем случае равенству (7) можно придать смысл, если переписать его в виде

$$\frac{d\nu}{dx}Lf(x)=\Delta f(x)-f(x)\frac{d\mu}{dx}$$

и понимать $d\nu/dx$, Δf и $d\mu/dx$ как обобщенные функции [1]. Доказывается,

---

[1] Произведение $f d\mu/dx=g$ определяется как функционал на финитных бесконечно дифференцируемых функциях, заданный формулой

$$g(\varphi)=\int\varphi f d\mu.$$

что если Lf определено этим равенством и функции f и Lf непрерывны в естественной топологии, то f принадлежит области определения характеристического оператора $\mathfrak{A}$ и $\mathfrak{A}f=Lf$. Отправляясь от этого описания оператора $\mathfrak{A}$ можно описать и инфинитезимальный оператор A (1).

Оператор L можно определить также следующим образом. Пусть f — функция точки и ψ — мера со знаком. Условимся писать $\psi=\Psi f$, если $d\psi/dx=-\Delta f$ (левые и правые части понимаются как обобщенные функции). Далее, пусть μ — некоторая мера. Будем писать $\psi=\mu f$, если $d\psi/d\mu=f$. Тогда

$$L=\frac{d}{dv}[\Psi+\mu].$$

Для винеровского процесса

$$L=\frac{d}{dx}\Psi.$$

7. 3. Из аналитических задач, связанных с обобщенным броуновским движением, мы рассмотрим две задачи: 1) изучение гармонических функций; 2) изучение переходной функции.

Доказывается, что функция f является гармонической для броуновского движения с убивающей мерой μ, тогда и только тогда, когда она непрерывна в естественной топологии и удовлетворяет уравнению

$$\Delta f=\frac{d\mu}{dx}f \tag{8}$$

(правая и левая части понимаются как обобщенные функции). Из общей теории гармонических функций, связанных с марковскими процессами, вытекают простые условия существования и единственности решения задачи Дирихле для уравнения (8) и явные вероятностные формулы, выражающие это решение. Пусть, в частности, мера μ сосредоточена на некоторой гладкой поверхности S и имеет там непрерывную плотность ϱ Тогда уравнение (8) сводится к уравнению Лапласа вне S и требованию, чтобы функция f испытывала при переходе через S скачок нормальной производной, равный $f\varrho$ (2).

Относительно переходной функции обобщенного броуновского движения известно пока еще немного. В одномерном случае $P(t,x,\Gamma)$ как функция от Γ имеет относительно меры скорости v плотность $p(t,x,y)$, симметричную относительно x и y. (Для процессов с нулевой убивающей мерой это было впервые доказано Вентцелем [52].) Верно ли аналогичное утверждение для многомерного случая, остается неизвестным.

---

(1) В случае, когда убивающая мера равна нулю, инфинитезимальные операторы описаны Маккином и Танака [33]. Общий случай рассмотрен в [16], [17].

(2) Это доказано Молчановым.

§ 8. Как устроен общий непрерывный строго марковский процесс?

8. 1. Одной из наиболее актуальных задач теории марковских процессов является выяснение строения наиболее общего непрерывного строго марковского процесса на топологическом многообразии E. Одномерный случай в настоящее время исследован полностью. Имеются некоторые подходы и к многомерному случаю.

Начнем с процессов, заданных на открытом интервале (r_1, r_2). Доказывается, что если из каждой точки интервала траектория может попасть с положительной вероятностью в любую другую точку, то процесс можно получить посредством монотонного непрерывного преобразования фазового пространства из некоторого обобщенного броуновского движения. Другими словами, его можно получить из винеровского процесса, последовательно выполняя три операции: образование подпроцесса, случайную замену времени и преобразование координат. Если процесс необрывающийся, то из этих трех операций нужны только две последние.

Сходный результат можно получить и относительно необрывающихся непрерывных строго марковских процессов на топологических многообразиях любого числа измерений. Однако, в общем случае исходным материалом для построения служит не только винеровский процесс, но и целый класс процессов, близкий к классу диффузионных процессов (мы называем эти процессы квазидиффузионными). К тому же сведение к квазидиффузионным процессам удается провести только локально и при некоторых ограничениях, которые нуждаются в дальнейшем анализе.

8. 2. Напомним, что непрерывный строго марковский процесс называется диффузионным, если его характеристический оператор определен на всех дважды непрерывно дифференцируемых функциях. Чтобы получить определение *квазидиффузионного* процесса, достаточно заменить здесь характеристический оператор на *квазихарактеристический* оператор, который определяется следующим образом. Пусть f и F — функции, заданные в некоторой окрестности точки x и пусть для каждой достаточно малой окрестности U этой точки

$$T_{\tau(U)} f(x) - f(x) = M_x \int_0^{\tau(U)} F(x_s) ds.$$

($\tau(U)$ — момент первого выхода из U.) Тогда мы говорим, что f принадлежит области определения квазихарактеристического оператора $\hat{\mathfrak{A}}$ в точке x и $\hat{\mathfrak{A}} f(x) = F(x)$. Если функция F непрерывна в точке x, то $\mathfrak{A} f(x)$ также равно $F(x)$, так что квазихарактеристический оператор совпадает с характеристическим (1).

(1) Аналогичным образом можно видоизменить определение инфинитезимального оператора. Условимся говорить, что функция f принадлежит области определения квазиинфинитезимального оператора $\hat{A}$ и $\hat{A} f = F$, если

$$T_t f - f = \int_0^t T_s F ds, \quad (t \geqslant 0).$$

8. 3. Рассмотрим теперь произвольный необрывающийся непрерывный строго марковский процесс (x_t, P_x) на многообразии E. Построение, сводящее такой процесс к квазидиффузионным процессам, основано на понятии гармонической функции и аддитивного функционала. Основная идея построения принадлежит Скороходу [43]. Скороход показал, что каждой паре гармонических функций f, g можно сопоставить непрерывный аддитивный функционал $\varphi_t^s(f,g)$, который определяется следующим образом (1). Разобьем отрезок $[s,t]$ на мелкие отрезки $[u_k, u_{k+1}]$, составим произведение приращений функций $f(x_u)$ и $g(x_u)$ на каждом из мелких отрезков. Сумма этих произведений при $\max|u_{k+1}-u_k|\to 0$ сходится в среднем квадратическом к $\varphi_t^s(f,g)$.

Пусть $f_1, \ldots, f_n$ — гармонические координаты вблизи точки x. Положим

$$\varphi_t^s = \sum_i \varphi_t^s(f_i, f_i).$$

Опираясь на теорему Радона-Никодима, можно показать, что

$$\varphi_t^s(f_i, f_j) = \int_s^t a_{ij}(x_u)\, d\varphi_u^0 .$$

Сделаем случайную замену времени, отвечающую функционалу φ_t^s (2). Мы получим квазидиффузионный процесс с производящим дифференциальным оператором

$$\sum a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}.$$

Если коэффициенты $a_{ij}(x)$ непрерывны, то этот процесс является диффузионным.

Условие существования гармонической системы координат является, как мы уже говорили, некоторым условием регулярности процесса. Таким образом исследование локального строения любых регулярных необрывающихся непрерывных строго марковских процессов сводится к изучению квазидиффузионных процессов.

Доказывается, что квазихарактеристический оператор является расширением оператора $\hat{A}$.

(1) Строго говоря, результат Скорохода относится к гармоническимфункциям, подчиненным дополнительному условию: при $t\downarrow 0$

$$\sup_x P_x\{|f(x_t)-f(x)|>\varepsilon\} = o(t)$$

для любого $\varepsilon>0$. Эту оговорку надо иметь в виду и в дальнейшем.

(2) Функционалы $\varphi_t^s(f_i, f_j)$ и φ_t^s определены до момента τ первого выхода из окрестности, в которой определены гармонические координаты. Случайная замена времени проводится после того, как исходный процесс оборван в момент τ.

§ 9. Неотрицательные гармонические функции и предельное поведение траекторий марковского процесса.

9. 1. В 1941 г. Мартин [34] исследовал множество всех неотрицательных решений уравнения Лапласа в произвольной области G евклидова пространства. Значение конструкций Мартина для теории вероятностей было выяснено в работах Дуба [3] и Ханта [29]. Дуб и Хант рассматривали гармонические функции, связанные с дискретными цепями Маркова. Случай диффузионных процессов был изучен недавно Шуром [41].

Мы начнем со следующей конкретной задачи. Пусть эллипсоид с закрепленным центром и фиксированным объемом случайно изменяется (поворачивается и растягивается). Предположим, что этот процесс является марковским, непрерывным и инвариантным относительно группы аффинных преобразований, не меняющих объема. Что будет происходить с эллипсоидом при $t\to\infty$?

Доказывается (см. [15]), что с вероятностью 1 главные направления эллипсоида стремятся к определенным предельным положениям и эллипсоид сплющивается; точнее, если $\varrho_1\geqslant\varrho_2\geqslant\varrho_3$ — длины его полуосей, то с вероятностью 1

$$\log\varrho_1\to\infty,\quad \frac{\log\varrho_2}{\log\varrho_1}\to 0,\quad \frac{\log\varrho_3}{\log\varrho_1}\to -1.$$

Множество всех эллипсоидов с закрепленным центром и фиксированным объемом представляет собой некоторое гладкое многообразие, и случайное движение, которое мы рассмотрели, является диффузионным процессом на этом многообразии. Пусть вообще на некотором гладком многообразии E задан диффузионный процесс с производящим дифференциальным оператором L. Оказывается, что вопрос о предельном поведении траектории при $t\to\infty$ тесно связан с вопросом о строении множества всех отвечающих процессу неотрицательных гармонических функций, т. е. решений уравнения $Lf=0$. Обозначим это множество через K и назовем функцию $f\in K$ минимальной, если ее нельзя разложить на сумму двух не пропорциональных друг другу функций из K.

В случае, когда L — оператор Лапласа в единичном шаре, то минимальные функции описываются формулой

$$f_\varrho(x)=\frac{1-|x|^2}{|x-\varrho|^3},$$

где $|\varrho|=1$ (через $|x|$ обозначается евклидова длина вектора x). Таким образом, минимальные функции находятся во взаимно однозначном соответствии с точками границы исходного шара. Оказывается, что и в общем случае можно погрузить фазовое пространство E в некоторый компакт таким образом, что минимальные функции оказываются во взаимно однозначном соответствии с некоторой частью границы E

(эта граница называется границей Мартина). Доказывается, что всякая неотрицательная гармоническая функция f записывается в виде интеграла

$$\int f_\varrho(x)\,\mu(d\varrho),$$

причем мера μ однозначно определяется по f. В частности, в предположении, что $L1=0$ (1), имеем

$$1=\int f_\varrho(x)\,\mu_0(d\varrho).$$

Доказывается, что траектория процесса входит с вероятностью 1 в определенную точку границы Мартина и вероятность того, что эта точка принадлежит множеству Γ, дается формулой

$$\int_\Gamma f_\varrho(x)\,\mu_0(d\varrho).$$

Таким образом, $f_\varrho(x)$ представляет собой плотность вероятности вхождения в точку ϱ для частицы, отправляющейся из точки x.

Сформулированный выше результат о предельном поведении диффундирующего эллипсоида получен именно таким путем. Выбирая какой-нибудь ортонормированный базис в евклидовом пространстве, можно описывать эллипсоиды положительно определенными матрицами. Оказывается, что минимальные функции представляют собой произведения степеней угловых миноров этой матрицы. Если мы хотим получить плотность вероятности того, что в пределе главные направления эллипсоида определяются ортонормированным базисом e_1, e_2, e_3, мы должны построить, отправляясь от базиса e_1, e_2, e_3, матрицу, соответствующую начальному положению эллипсоида, и взять произведение угловых миноров этой матрицы в степени -2. (Все сказанное распространяется на эллипсоиды в комплексном евклидовом пространстве любого числа измерений (см. [15]). Аналогичные результаты можно получить и для других симметрических пространств отрицательной кривизны.

Множество минимальных неотрицательных решений эллиптического дифференциального уравнения $Lf=0$ еще мало исследовано и здесь имеется много интересных нерешенных задач. Мы сформулируем две такие задачи.

Первая задача: при каких условиях на дифференциальный оператор L можно ввести в множество минимальных функций структуру гладкого многообразия так, чтобы $f_\varrho(x)$ являлась гладкой функцией от ϱ и x?

Вторая задача касается связей между локальной геометрией полного риманова многообразия и строением границы Мартина этого многообразия. (В качестве оператора L рассматривается дифференциальный оператор Бельтрами, определяемый римановой метрикой.) Во всех примерах полных односвязных многообразий отрицательной кривизны,

---

(1) Это предположение равносильно требованию, чтобы траектория процесса выходила, не обрываясь, из любого компакта.

для которых подсчитана размерность множества минимальных функций, эта размерность отличается на единицу от размерности многообразия. Верно ли, что и в общем случае отрицательность кривизны и односвязность влекут за собой определенное богатство множества минимальных функций?

ЛИТЕРАТУРА

[1]. Blumenthal, R. M., Getoor, R. K., & McKean, H. P., Jr., Markov processes with identical hitting distributions. *Illinois J. Math.*, 8 (1962), 402–421.

[2]. Doob, J. L., Semimartingales and subharmonic functions. *Trans. Amer. Math. Soc.*, 77 (1954), 86–121.

[3]. —— Discrete potential theory and boundaries. *J. Math. Mech.*, 8 (1959), 433–458.

[4]. Дынкин, Е. Б., Функционалы от траекторий марковских случайных процессов. *Доклады АН СССР*, 104 (1955), 691–694.

[5]. —— Бесконечно малые операторы марковских случайных процессов. *Доклады АН СССР*, 105 (1955), 206–209.

[6]. —— Непрерывные одномерные марковские процессы. *Доклады АН СССР*, 105 (1955), 405–408.

[7]. —— Марковские процессы и полугруппы операторов. *Теор. вер. и ее прим.*, 1 (1956), 25–37.

[8]. —— Инфинитезимальные операторы марковских процессов. *Теор. вер. и ее прим.*, 1 (1956), 38–60.

[9]. —— Одномерные непрерывные строго марковские процессы. *Теор. вер. и ее прим.*, 4 (1959), 3–54.

[10]. —— Естественная топология и эксцессивные функции, связанные с марковским процессом. *Доклады АН СССР*, 127 (1959), 17–19.

[11]. —— *Основания теории марковских процессов.* Москва, 1959.

[12]. —— Марковские процессы и связанные с ними задачи анализа. *Успехи матем. наук*, 15: 2 (92) (1960), 3–24.

[13]. —— Аддитивные функционалы от винеровского процесса, определяемые стохастическими интегралами. *Теор. вер. и ее прим,*. 5 (1960), 441–452.

[14]. Dynkin, E. B., Transformations of Markov processes connected with additive functionals. *Proc. Fourth Berkeley Symp. on Math. Stat. and Prob.*, 2, 117–142. Berkeley, 1961.

[15]. —— Неотрицательные собственные функции оператора Лапласа-Бельтрами и броуновское движение в некоторых симметрических пространствах. *Доклады АН СССР*, 141 (1961), 288–291.

[16]. —— Броуновское движение с убивающей мерой μ и мерой скорости ν. *Доклады АН СССР*, 144 (1962), 483–486.

[17]. —— *Марковские процессы.* Москва, 1963.

[18]. Дынкин, Е. Б., и Юшкевич, А. А., Строго марковские процессы. *Теор. вер. и ее прим* ., 1 (1956), 149–155.

[19]. Feller, W., The parabolic differential equations and the semi-groups of transformations. *Ann. Math.*, 55 (1952), 468–519.

[20]. —— Generalized second order differential operators and their lateral conditions. *Illinois J. Math.*, 1 (1957), 459–504.

[21]. Фрейдлин, М. И., Диффузионные процессы с отражением и задача с

косой производной на многообразии с краем. *Теор. вер. и ее прим.*, 8 (1963), 80–88.

[22]. —— Задача Дирихле для уравнения с малым параметром и разрывными коэффициентами. *Доклады АН СССР*, 144 (1962), 501–504.

[23]. Гирсанов, И. В., О некоторых топологиях, связанных с марковским процессом. *Доклады АН СССР*, 129 (1959), 488–491.

[24]. Хасьминский, Р. З., Диффузионные процессы и эллиптические дифференциальные уравнения, вырождающиеся на границе области. *Теор. вер. и ее прим.*, 3 (1958), 430–451.

[25]. —— О положительных решениях уравнения $\mathfrak{A}u + Vu = 0$. *Теор. вер. и ее прим.* 4 (1959), 332–341.

[26]. —— Эргодические свойства возвратных диффузионных процессов и стабилизация решений задачи Коши для параболических уравнений. *Теор. вер. и ее прим.*, 5 (1960), 196–214.

[27]. —— О принципе усреднения для параболических и эллиптических дифференциальных уравнений и марковских процессов с малой диффузией. *Теор. вер. и ее прим.*, 8 (1963), 3–25.

[28]. Hunt, G. A., Markoff processes and potentials. *Illinois J. Math.*, 1 (1957), 44–93; 1 (1957), 316–369; 2 (1958), 151–213.

[29]. —— Markoff chains and Martin boundaries. *Illinois J. Math.*, 4 (1960), 313–340.

[30]. Itô, K., Stochastic differential equations in a differentiable manifold. I. *Nagoya Math. J.*, 1 (1950), 35–47; II. *Mem. Coll. Sci., Univ. Kyoto, Ser. A, 28, Mathematics*, 1 (1953), 81–85.

[31]. —— On stochastic differential equations. *Mem. Amer. Math. Soc.*, 4 (1951).

[32]. Itô, K., & McKean, H. P., Jr. *Diffusion theory*. (To appear.)

[33]. McKean, H. P., Jr., & Tanaka, H., Additive functionals of the Brownian path. *Mem. Coll. Sci., Univ. Kyoto, Ser. A, Mathematics*, 33 (1961), 479–506.

[34]. Martin, R. S., Minimal positive harmonic functions. *Trans. Amer. Math. Soc.*, 49 (1941), 137–172.

[35]. Meyer, P. A., Fonctionelles multiplicatives et additives de Markov. *Ann. Inst. Fourier, Grenoble*, 12 (1962), 125–230.

[36]. Pogorzelski, W., Étude de la solution fondamentale de l'équation parabolique. *Ricerche di Mat., Napoli*, 5 (1956), 25–37.

[37]. Погожельский, В., Исследование интегралов параболического уравнения и краевых задач в неограниченной области. *Матем. сборн.* 47 (1959), 397–430.

[38]. Ray, D., Stationary Markov processes with continuous paths. *Trans. Amer. Math. Soc.*, 82 (1956), 452–493.

[39]. Шур, М. Г., Непрерывные аддитивные функционалы от марковских процессов и эксцессивные функции. *Доклады АН СССР*, 137 (1961), 800–803.

[40]. —— Эксцессивные функции и аддитивные функционалы от марковских процессов. *Доклады АН СССР*, 143 (1962), 293–296.

[41]. —— Граница Мартина для линейного эллиптического оператора второго порядка. *Известия АН СССР, сер. матем.*, 27 (1963), 45–60.

[42]. Скороход, А. В., Аддитивные функционалы от процесса брауновского движения. *Теор. вер. и ее прим.*, 6 (1961), 430–439.

[43]. —— Об однородных непрерывных марковских процессах, являющихся мартингалами. *Теор. вер. и ее прим.*, 8 (1963).

[44]. Ueno, T., The diffusion satisfying Wentzell's boundary conditions and the Markov processes on the boundary. *Proc. Japan Acad.*, 36 (1960), 533–538, 625–629.

[45]. Волконский, В. А., Случайная замена времени в строго марковских процессах. *Теор. вер. и ее прим.*, 3 (1958), 332–350.

[46]. —— Аддитивные функционалы от марковских процессов. *Труды Москов. матем. общества*, 9 (1960), 143–189.

[47]. Вентцель, А. Д., О граничных условиях для многомерных диффузионных процессов. *Теор. вер. и ее прим.*, 4 (1959), 172–185.

[48]. —— Общие граничные задачи, связанные с диффузионными процессами. *Успехи матем. наук*, 15 (1960), 202–204.

[49]. —— Неотрицательные аддитивные функционалы от марковских процессов. *Доклады АН СССР*, 137 (1961), 17–20.

[50]. —— Аддитивные функционалы от многомерного винеровского процесса. *Доклады АН СССР*, 139 (1961), 13–16.

[51]. —— О непрерывных аддитивных функционалах от многомерного винеровского процесса. *Доклады АН СССР*, 142 (1962), 1223–1226.

[52]. —— Об абсолютной непрерывности переходных вероятностей одномерного диффузионного процесса. *Теор. вер. и ее прим.*, 6 (1961), 439–446.

HOMOTOPY AND COHOMOLOGY THEORY

By BENO ECKMANN

1. Introduction

Algebraic topology has become a quite complex discipline. This is not only due to many new ramifications and to highly developed, refined and powerful algebraic techniques, but also to the topic itself. Its aim is to solve geometrical problems by algebraic computation; but these problems come from very different sources ranging from point-set theory in Euclidean space to differential geometry and Lie group theory and to algebraic geometry and complex analysis. The algebraic-topological methods have in many cases been suggested by the respective fields, and they have, conversely, strongly influenced them and contributed to their recent development.

Nevertheless there has been in recent years a trend to unification, which clarifies the common ideas and is helpful for applications. At its origin is, on one hand, the functorial formulation due to Eilenberg and MacLane and the axiomatic approach to homology due to Eilenberg and Steenrod; and, on the other hand, the successful efforts made by many authors in homotopy and fiber bundle theory, including operations, manifold theory, extraordinary cohomology etc., where it would be difficult to list all the names. It appeared more and more clear that there is a common framework for the various concepts and methods, simple at least in its basic ideas. It is this simplicity I wish to emphasize in the following, without attempting any degree of completeness.

The unified approach I would like to present in this expository lecture is based on the simple and intuitive geometrical concept of *homotopy*. Thus our starting point for the algebraization is the set $\Pi(X, Y)$ of homotopy classes of maps of the space X into the space Y. The following conventions are valid throughout: All spaces are assumed to be provided with a base-point written o, all maps $f\colon X \to Y$ carry base-points into base-points, and all homotopies (i.e., continuous one-parameter families of maps $X \to Y$) respect base-points for all values of the deformation parameter; in other words, the image of the base-point remains fixed during the deformation. The set $\Pi(X, Y)$ contains a distinguished element 0, namely the homotopy class of the constant map $X \to o \in Y$, the maps in that class being called "contractible" or null-homotopic. A map $h\colon Y \to Y'$ induces, by composition, a map h_* of the set $\Pi(X, Y)$ into $\Pi(X, Y')$, with $h_*(0) = 0$; and a map $g\colon X' \to X$ induces a map g^* of $\Pi(X, Y)$ into $\Pi(X', Y)$, with $g^*(0) = 0$. The set $\Pi(X, Y)$ is, of course, a topological invariant of both X and Y; more precisely, it is a functor from based spaces to based sets, covariant in Y and contravariant in X, the map h_* induced by a homeomorphism $h\colon Y \to Y'$ (or g^* induced by a homeomorphism $g\colon X' \to X$) being an equivalence of sets. In fact $\Pi(X, Y)$ depends only on the homotopy type of X and of Y: if $h\colon Y \to Y'$ is a homotopy equivalence (i.e., if there is a map $h'\colon Y' \to Y$ such that hh' and $h'h$ are homotopic to the respective identities), then h_* is an equivalence of the sets $\Pi(X, Y)$ and $\Pi(X, Y')$; and similarly for a homotopy equivalence $g\colon X' \to X$.

Our starting point thus seems to make use of only a weak part of the structure present in the spaces involved, and to eliminate all those aspects which are not homotopy invariant. The situation, however, is as follows: In general neither the problems nor the results are homotopy invariant; but whenever it comes to actual algebraization and computation, the construction depends only on the homotopy type—not of the original space, manifold variety, etc., but of other spaces associated with them by procedures (which need not be homotopy invariant, nor even of topological nature) which use as much of the given structures as possible.

2. Two-space homotopy groups

For a fixed space Y the set $\Pi(X, Y)$ is a contravariant functor of the space X which will be used to investigate X; and vice versa, for a fixed X, a covariant functor of Y used to investigate Y. The structure "set with distinguished element 0" is, of course, not good enough for algebraic purposes, and one wants to give this set a *natural group structure*, with 0 being the neutral element of the group. "Natural" means that the structure in $\Pi(X, Y)$, for fixed Y, turns $\Pi(X, Y)$ into a functor from spaces to groups, the induced maps g^* being homomorphisms; and similarly for fixed X. We do not go here into the general question of finding all test-spaces Y or X respectively for which this is possible (cf. [4, 5]), but simply list the two cases which will be used in the following.

(*a*) For an arbitrary space B, let ΩB be the space of loops in B beginning and ending at o, with the constant loop as base-point; ΩB is provided with a multiplication (a map $\Omega B \times \Omega B \to \Omega B$), namely the usual composition of loops, fulfilling the group axioms up to homotopy. Two maps $f, f' : X \to \Omega B$ can be multiplied by multiplying the images $f(x)$ and $f'(x)$ in ΩB, and this establishes in $\Pi(X, \Omega B)$ a natural group structure.

(*b*) For a space A, let ΣA be the suspension of A, i.e., the double cone with basis A; more precisely, ΣA is the product space of A with the unit interval $I = \{0 \leqslant t \leqslant 1\}$, where the union $A \times (0) \cup A \times (1) \cup o \times I$ is identified to single point, the base-point of ΣA. There is a "comultiplication" μ in ΣA, i.e. a map of ΣA into $\Sigma A \vee \Sigma A$ (union with identified base-points); it essentially consists in pinching $A \times (\frac{1}{2})$ to a point. For an arbitrary space Y, μ can be used to define a multiplication for maps $f, f' : \Sigma A \to Y$, as follows: Let $F : \Sigma A \vee \Sigma A \to Y$ be the map given by f on the first copy ΣA and by f' on the second, and define $F\mu : \Sigma A \to Y$ to be the product of f and f'. It is easily checked that this establishes a group structure in $\Pi(\Sigma A, Y)$, natural with respect to Y.

Moreover, the usual identification of maps from $A \times I$ to B with maps of A into the mapping space B^{I} immediately yields a natural equivalence between $\Pi(\Sigma A, B)$ and $\Pi(A, \Omega B)$ which is a group isomorphism. We thus identify the groups $\Pi(\Sigma A, B) = \Pi(A, \Omega B)$ and define *two-space homotopy groups* $\Pi_n(A, B), n \geqslant 0$, by

$$\Pi_n(A, B) = \Pi(\Sigma^n A, B) = \Pi(\Sigma^{n-k} A, \Omega^k B), \quad 0 \leqslant k \leqslant n,$$

where Σ^n is the n-fold suspension and Ω^n the n-fold loop-space. For $n \geqslant 2$, we can use in $\Pi_n(A, B) = \Pi(\Sigma^{n-1} A, \Omega B)$ either the multiplication in ΩB or the comultiplication in $\Sigma^{n-1} A$ for defining the group structure; by a simple

and general argument the two group structures coincide and are Abelian (cf. [4, 5]). We thus obtain, for $n \geqslant 2$, functors $\Pi_n(A, B)$ from (based) spaces to Abelian groups, covariant in B and contravariant in A, with homotopic maps inducing the same homomorphisms. We sometimes write $\Pi_*(A, B)$ for the direct sum of all $\Pi_n(A, B)$, $n \geqslant 2$. If A is a sphere S_{m-n}, then $\Pi_n(A, B) = \Pi(\Sigma^n S_{m-n}, B) = \Pi(S_m, B) = \pi_m(B)$ is the usual mth Hurewicz homotopy group of B; the well-known and important algebraic feature of these groups is expressed by an exact sequence. In the next section we show more generally that there are two dual exact sequences for the $\Pi_n(A, B)$, one for a fixed A and one for a fixed B. The "duality" appearing here and in other instances consists in exchanging the role of A and B, i.e., in reversing the directions of all maps; it can be given an abstract foundation, cf. [5] and [6].

3. Exact sequences

For a fixed space A, we consider the group $\Pi_n(A, B)$ and, for a given map $\beta: B \to B'$, the induced homomorphism $\beta_*: \Pi_n(A, B) \to \Pi_n(A, B'), n \geqslant 2$. In the usual manner, the deviation of these β_* from being isomorphisms is measured by "relative" groups; this means that we can define groups $\Pi_n(A, \beta)$ attached to the map β (and to A) and two homomorphisms J: $\Pi_n(A, B') \to \Pi_n(A, \beta)$ and $\partial: \Pi_n(A, \beta) \to \Pi_{n-1}(A, B)$ such that the sequence

$$\ldots \to \Pi_n(A, B) \xrightarrow{\beta} \Pi_n(A, B') \xrightarrow{J} \Pi_n(A, \beta) \xrightarrow{\partial} \Pi_{n-1}(A, B) \to \ldots$$

is exact. We do not give here the full details of these definitions, but only some indications to show that everything is obtained in a quite elementary geometrical fashion.

In order to define $\Pi_n(A, \beta)$, we have to consider maps $f: \alpha \to \beta$ "of the map α into the map β". We write CX for the cone over X (i.e., the space $X \times I$ with $X \times (1) \cup o \times I$ identified to a single point o) and take for α the inclusion map $\Sigma^{n-1}A \to C(\Sigma^{n-1}A)$ of the basis into the cone. A map $f: \alpha \to \beta$ is then given by two "components" $f_0: \Sigma^{n-1}A \to B$ and $f_1: C(\Sigma^{n-1}A) \to B'$ subject to the condition that the diagram

$$\begin{array}{ccc} \Sigma^{n-1}A & \xrightarrow{f_0} & B \\ \alpha\downarrow & & \downarrow\beta \\ C(\Sigma^{n-1}A) & \xrightarrow[f_1]{} & B' \end{array}$$

be commutative. By a homotopy of f we understand a pair of homotopies of f_0 and f_1 respectively such that the above diagram is commutative for all values of the deformation parameter. The homotopy classes of these maps $f: \alpha \to \beta$ are the elements of the group $\Pi_n(A, \beta)$; the group structure is again defined with the help of the suspension Σ (for $n \geqslant 2$; it is Abelian for $n \geqslant 3$). If B' consists of a single point o, one has $\Pi_n(A, \beta) \cong \Pi_{n-1}(A, B)$ by an obvious isomorphism (omitting the component f_1). If B consists of a single point, one has $\Pi_n(A, \beta) \cong \Pi_n(A, B')$, since the space $C(\Sigma^{n-1}A)/\Sigma^{n-1}A$, obtained from the cone by identifying its basis to a single point o, is homeomorphic to $\Sigma^n A$.

In order to define J and ∂, we consider the commutative diagram

$$\begin{array}{ccccccc} & & & & & & o \\ & & & & & & \downarrow\gamma \\ o & \rightarrow & o & \rightarrow & B & \xrightarrow{1} & B \\ \gamma\downarrow & & \downarrow\gamma' & & \downarrow\beta & & \downarrow\tilde{\gamma} \\ B & \xrightarrow{\beta} & B' & \xrightarrow{1} & B' & \rightarrow & o \end{array}$$

where β is the given map $B \rightarrow B'$, γ and γ' are the embeddings of the base-point into B and B' respectively, and $\tilde{\gamma}$ is the trivial map $B \rightarrow o$. The horizontal maps define three maps: $\gamma \rightarrow \gamma'$, $\gamma' \rightarrow \beta$ and $\beta \rightarrow \tilde{\gamma}$, with induced homomorphisms $\Pi_n(A,\gamma) \rightarrow \Pi_n(A,\gamma') \rightarrow \Pi_n(A,\beta) \rightarrow \Pi_n(A,\tilde{\gamma}) \cong \Pi_{n-1}(A,\gamma)$; replacing the first group by $\Pi_n(A,B)$, the second by $\Pi_n(A,B)$ and the fourth by $\Pi_{n-1}(A,B)$ we get the three homomorphisms β_*, J and ∂ of the sequence above. From these definitions, exactness is easily obtained.

[If we take for A a sphere S_{m-n}, $\Pi_n(A,B) = \pi_m(B)$ is the mth Hurewicz homotopy group of B; if $\beta: B \rightarrow B'$ is an inclusion map of a subspace $B \subset B'$, $\Pi_n(A,\beta)$ is the relative Hurewicz homotopy group $\pi_m(B' \bmod B)$, and the sequence is the classical relative Hurewicz homotopy sequence, which thus appears as the first natural case of algebraization. Obviously the assumption that β be an inclusion map is irrelevant.]

From the above considerations it is clear that the groups appearing in the exact sequence are in fact special cases of relative groups, and the whole argument generalizes easily to arbitrary relative groups, as follows. In the diagram above we replace the two first vertical maps by non-trivial maps $\gamma: B_0 \rightarrow B$, $\gamma': B_0 \rightarrow B'$ of a space B_0 into B and B':

$$\begin{array}{ccccccc} & & & & & & B_0 \\ & & & & & & \downarrow\gamma \\ B_0 & \xrightarrow{1} & B_0 & \xrightarrow{\gamma} & B & \xrightarrow{1} & B \\ \gamma\downarrow & & \downarrow\gamma' & & \downarrow\beta & & \downarrow\tilde{\gamma} \\ B & \xrightarrow[\beta]{} & B' & \xrightarrow[1]{} & B' & \rightarrow & o \end{array}$$

Commutativity of the diagram then requires that $\gamma' = \beta\gamma$. The group $\Pi_n(A,\tilde{\gamma}) = \Pi_{n-1}(A,B)$ can no longer be identified with the group $\Pi_{n-1}(A,\gamma)$, but there is an obvious homomorphism η into it. Thus the three homomorphisms induced by the vertical maps (the third one composed with η) yield a sequence

$$\ldots \rightarrow \Pi_n(A,\gamma) \rightarrow \Pi_n(A,\beta\gamma) \rightarrow \Pi_n(A,\beta) \rightarrow \Pi_{n-1}(A,\gamma) \rightarrow \ldots.$$

of relative groups for the maps γ, β and $\beta\gamma: B_0 \rightarrow B \rightarrow B'$. As before the sequence is easily seen to be exact. It is called the *triple sequence* for the "triple" of spaces $B_0 \xrightarrow{\gamma} B \xrightarrow{\beta} B'$, *relating the three relative groups of* γ, *of* β, *and of the composition* $\beta\gamma$, for a fixed A. It establishes a precise relation between $\Pi_*(A,\gamma)$ and $\Pi_*(A,\beta)$ on one hand, and $\Pi_*(A,\beta\gamma)$ on the other; this relation does not determine $\Pi_*(A,\beta\gamma)$ but expresses it in a well-known fashion "up to a group extension". The ordinary exact sequence is a special case of the triple sequence; the latter has the advantage of making use of the additional structure present in the category of maps between spaces, namely the composition of maps.

The next step then is obviously the take a composition of more than two maps $\beta = \beta_1 \beta_2 \ldots \beta_r$ (for some integer $r \geqslant 2$; the infinite case could be handled similarly, but is omitted here for simplicity):

$$B_r \xrightarrow{\beta_r} B_{r-1} \to \ldots \xrightarrow{\beta_2} B_1 \xrightarrow{\beta_1} B$$

and to ask for the relationship between the $\Pi_*(A, \beta_i), i = 1, \ldots, r$ on one hand and $\Pi_*(A, \beta)$ on the other. The answer is clearly given by a *spectral sequence*: one starts from the direct sum $\sum_{i=1}^{r} \Pi_*(A, \beta_i)$ and amalgamates the triple sequences corresponding to the successive compositions into exact couples; the spectral sequence thus obtained (cf. Eckmann–Huber [6]) begins with $E_1 = \sum_{i=1}^{r} \Pi_*(A, \beta_i)$ and ends with $\Pi_*(A, \beta)$ in the usual sense that E_∞ is the graded group of $\Pi_*(A, \beta)$ for an obvious filtration (given by the factorization $\beta = \beta_1 \beta_2 \ldots \beta_r$). The differentials of the spectral sequence are obtained from the homomorphisms appearing in the various triple sequences.

Just a few words about the dual situation: In the groups $\Pi_n(A, B)$, a fixed test-space B is considered; for a map $\alpha: A \to A'$ with induced homomorphisms $\alpha^*: \Pi_n(A', B) \to \Pi_n(A, B)$ a relative group $\Pi_n(\alpha, B)$ is defined and homomorphisms $J: \Pi_n(A, B) \to \Pi_n(\alpha, B)$, $\partial: \Pi_n(\alpha, B) \to \Pi_{n-1}(A', B)$ such that the sequence

$$\ldots \to \Pi_n(A', B) \xrightarrow{\alpha^*} \Pi_n(A, B) \xrightarrow{J} \Pi_n(\alpha, B) \xrightarrow{\partial} \Pi_{n-1}(A', B) \to \ldots$$

is exact. The elements of $\Pi_n(\alpha, B)$ are the homotopy classes of maps of α into β, where β designates the projection map $E\Omega^{n-1}B \to \Omega^{n-1}B$ (EX denoting the space of all open paths in X ending at o, with projection $\beta: EX \to X$, which is a fiber space over X with fiber $\beta^{-1}(o) = \Omega X$):

$$\begin{array}{ccc} A & \to & E\Omega^{n-1}B \\ \alpha\downarrow & & \downarrow\beta \\ A' & \to & \Omega^{n-1}B \end{array}$$

More generally, for the composition $\gamma\alpha$ of two maps $A \xrightarrow{\alpha} A' \xrightarrow{\gamma} A_0$ one has an exact *triple sequence*

$$\ldots \to \Pi_n(\gamma, B) \to \Pi_n(\gamma\alpha, B) \to \Pi_n(\alpha, B) \to \Pi_{n-1}(\gamma, B) \to \ldots,$$

and for the composition of r maps $\alpha = \alpha_r \alpha_{r-1} \ldots \alpha_1$ a *spectral sequence* beginning with the direct sum $E_1 = \sum_{i=1}^{r} \Pi_*(\alpha_i, B)$ and ending with $\Pi_*(\alpha, B)$ for a filtration corresponding to the factorization.

The two dual results, incidentally, need not be established separately, but can be obtained from a general categorical argument using semi-simplicial constructions (cf. P. J. Huber [7]).

Before going into the applications of this general algebraic framework, all based upon a special choice of the test-spaces A or B respectively, we describe a further general feature known under the name of "excision". It again manifests itself in two perfectly dual forms.

4. Excision

A map $\beta: E \to B$ is called a *fiber map*, with fiber $F = \beta^{-1}(o)$, if the *covering homotopy property* holds for arbitary maps $f: X \to B$; i.e., if f is factored through

β as $f=\beta f'$, then any homotopy of f can be factored through β. For the Hurewicz homotopy groups it is then a well-known fact that $\pi_n(\nu)$ is naturally isomorphic to $\pi_n(B)$, where ν denotes the inclusion $F\subset E$; this is easily seen to be equivalent to the isomorphism $\pi_n(\beta)\cong\pi_{n-1}(F)$, and in this form a general statement can be formulated, expressing relative groups $\Pi_n(A,\beta)$ of a fiber map β by "absolute" groups of the fiber F: The commutative diagram

$$\begin{array}{ccc} F & \xrightarrow{\nu} & E \\ {\scriptstyle\beta_0}\downarrow & & \downarrow{\scriptstyle\beta} \\ o & \rightarrow & B \end{array}$$

considered as a map of β_0 into β induces a homomorphism $\Pi_n(A,\beta_0)\rightarrow \Pi_n(A,\beta)$, called excision; from the fiber map property it follows easily that this excision homomorphism is an isomorphism. The left-hand side being $\cong\Pi_{n-1}(A,F)$, one has a natural isomorphism $\Pi_{n-1}(A,F)\cong\Pi_n(A,\beta)$. [Equivalently one might say that for a fiber map β the group $\Pi_n(A,\beta)$ remains unchanged if everything but the fiber is removed from E; i.e., if β is replaced by β_0. The exact sequence then gives the above isomorphism.]

Dually a map $\alpha:A\rightarrow E$ is called a *cofiber map*, with cofiber $F=E/\alpha(A)$ (the image in E identified to a single point o), if the *homotopy extension property* holds for arbitrary maps $f:A\rightarrow X$; i.e., if f is factored through α as $f=f'\alpha$, then any homotopy of f can be factored through α. For an inclusion $A\subset E$ this is the ordinary homotopy extension property. The diagram

$$\begin{array}{ccc} A & \rightarrow & o \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha_0} \\ E & \rightarrow & F \end{array}$$

then induces an excision isomorphism $\Pi_n(\alpha_0,B)\rightarrow\Pi_n(\alpha,B)$, where the right-hand side is $\cong\Pi_{n-1}(F,B)$. Thus the relative group $\Pi_n(\alpha,B)$ of a cofibration is naturally isomorphic to the "absolute" group $\Pi_{n-1}(F,B)$ of the cofiber. In other words, the relative group of a cofibration only depends on the cofiber $F=E/\alpha(A)$ and remains unchanged if the interior of $\alpha(A)$ is removed.

These two excision isomorphisms for fiberings and cofiberings, converting relative groups into absolute ones, prove in certain cases useful for computation.

5. Spectra. Cohomology

From the general set-up the groups normally used in algebraic topology are obtained by special choices of the test-spaces A and B respectively. As said before, taking for A a sphere (e.g., S_0) yields the Hurewicz homotopy groups with their exact sequences, triple sequences and spectral sequences, and with the excision property for fiberings. In the dual situation, cohomology theory is obtained by using test-spaces B belonging to a "spectrum", as explained below (there is a certain lack of duality appearing at this stage, which will be discussed later).

An *Ω-spectrum* $\mathfrak{B}$, or in short a spectrum, is a sequence of spaces B_k given for all integers k, together with maps $w_k:B_k\rightarrow\Omega B_{k+1}$ which are *homotopy*

equivalences for all k. Given such a spectrum $\mathfrak{B}=(B_k, w_k)$, there is an isomorphism (given by the respective w_k) between the groups $\Pi_n(A, B_{m+n}) = \Pi(A, \Omega^n B_{m+n})$ and $\Pi(A, B_m)$, for any $n>0$. Thus the two groups can be identified; thus $\Pi_n(A, B_{m+n}) = \Pi(A, B_m)$ is independent of $n>0$ and we denote this Abelian group by $h^m(A)$, m being an arbitrary integer. Similarly, relative groups $h^m(\alpha)$ for a map $\alpha: A \to A'$ are defined by $h^m(\alpha) = \Pi_{n+1}(\alpha, B_{m+n}) = \Pi_1(\alpha, B_m)$ for all m; for $\alpha^0: o \to A$ one has $h^m(\alpha^0) = h^m(A)$, for $\alpha_0: A \to o$, one has $h^m(\alpha_0) = h^{m-1}(A)$. The exact sequence of the groups Π_n

$$\ldots \to \Pi_n(A', B_{m+n}) \xrightarrow{\alpha^*} \Pi_n(A, B_{m+n}) \xrightarrow{J} \Pi_n(\alpha, B_{m+n}) \xrightarrow{\partial} \Pi_{n-1}(A', B_{m+n}) \to \ldots$$

then translates into an exact sequence

$$(S) \qquad \ldots \to h^m(A') \xrightarrow{\alpha^*} h^m(A) \xrightarrow{J} h^{m+1}(\alpha) \xrightarrow{\partial} h^{m+1}(A') \to \ldots$$

which looks like a cohomology sequence, where the upper indices correspond to dimensions and where J is usually denoted by δ. The excision property for cofibrations $\Pi_{n+1}(\alpha, B_{n+m}) \cong \Pi_n(F, B_{m+n})$, F being the cofiber of α, translates into the isomorphism $h^m(F) \cong h^m(\alpha)$, again in agreement with cohomology excision; the isomorphism is induced by

$$\begin{array}{ccc} A & \to & o \\ \alpha\downarrow & & \downarrow\varphi \\ A' & \to & F \end{array}$$

considered as a map of α into φ. A third cohomology property to be mentioned here, obvious from our homotopical definitions, is the fact that homotopic maps α induce the same homomorphisms $\alpha^*: h^m(A') \to h^m(A)$.

To sum up: We have obtained a sequence of contravariant functors $h^m(\alpha)$ from the category of based maps α of spaces, to the category of Abelian groups, defined for all integers m and such that for any space A the two maps $\alpha^\circ: o \to A$ and $\alpha_0: A \to o$ are related by a natural "connecting" isomorphism $h^m(\alpha^0) \cong h^{m+1}(\alpha_0)$; this group is written $h^m(A)$, defining a contravariant functor from spaces to Abelian groups. These functors have the three properties (I) Exactness, (II) Excision, (III) Homotopy; (I) means that the sequence (S) induced by the diagram

$$\begin{array}{ccccccc} o & \to & A & \xrightarrow{1} & A & \xrightarrow{\alpha} & A' \\ \downarrow & & \downarrow\alpha & & \downarrow & & \downarrow \\ A' & \xrightarrow{1} & A' & \to & o & \to & o \end{array}$$

and the "connecting isomorphism" is exact; (II) means that for a cofibration α the group $h^m(\alpha)$ depends only on the cofiber, and (III) that homotopic maps induce the same homomorphism. Any system of functors with these properties is called a *general cohomology theory*. We thus summarize the above considerations by saying that, *given a spectrum* $\mathfrak{B}=(B_m, w_m)$, *our homotopical definitions yield a general cohomology theory* $h^m(\alpha)$. We will limit ourselves here entirely to such homotopically defined cohomology theories; by a recent representation theorem of R. Brown [2] this is, to a large extent not a real restriction: the theorem states that (under certain

countability assumptions) any abstractly given cohomology theory can be realized homotopically by means of a spectrum $\mathfrak{B}$.

Here the following remarks should be added for completeness or in view of later application.

(1) We have restricted ourselves to "reduced" cohomology theories h^*, with $h^m=0$ for a one-point space and with groups and homomorphisms defined for spaces with base-points and based maps only. "Unreduced" theories h_+^* for free spaces[1] can be obtained from these reduced ones (and vice versa) by standard procedures (cf., e.g., [8]), and there are no essentially different features. In particular, if we write A^+ for the topological sum of A with an extra point which is taken as base-point of A^+, the groups $h_+^m(A)$ of h_+^* are given by $h^m(A^+)$; the map $A^+ \to A$ which is the identity on A induces a monomorphism $h^m(A) \to h_+^m(A)$.

(2) The exactness property (I) includes, in fact, the triple sequence for the composition of two maps; and consequently also the *spectral sequence* for the composition of $r \geqslant 2$ maps $\alpha = \alpha_r \alpha_{r-1} \dots \alpha_1$, with $E_1 = \sum_{i=1}^r h^*(\alpha_i)$ and E_∞ equal to the graded group associated with the filtration of $h^*(\alpha)$ determined by the factorization.

(3) From the definitions it follows immediately that $\Pi(\Sigma A, B_{m+1}) \cong \Pi(A, \Omega B_{m+1}) = \Pi(A, B_m)$, i.e., $h^{m+1}(\Sigma A) \cong h^m(A)$. Thus, for example, the cohomology h^* of the k-sphere S_k can be expressed by $h^*(S_0)$, with $h^m(S_k) \cong h^{m-k}(S_0)$ for all m.

(4) The "dimension axiom" of cohomology requires that $h^m(S_0)=0$ for all $m \neq 0$, or equivalently, that for any $k \geqslant 0$ $h^m(S_k)=0$ for all $m \neq k$. We note that $h^m(S_k) = \Pi(S_k, B_m) = \pi_k(B_m)$, and there may very well be spectra for which *this dimension axiom does not hold*. In order that this axiom holds, it is necessary and sufficient that, for each integer m, $\pi_k(B_m)=0$ for $k \geqslant 0$, $k \neq m$; this means for $m \geqslant 0$ that B_m is an Eilenberg–MacLane space $K(G_m, m)$ for the dimension m and the Abelian group $G_m = \pi_m(B_m)$, and for $m<0$ that B_m is (homotopy equivalent to) a one-point space o. Moreover the homotopy equivalence $w_m: B_m \to \Omega B_{m+1}$ shows that $G_m \cong G_{m+1}, m \geqslant 0$. Thus the *cohomology with respect to a spectrum $\mathfrak{B}$ fulfills the dimension axiom if and only if $\mathfrak{B}$ is the Eilenberg-MacLane spectrum*: $B_m = K(G, m)$ for $m \geqslant 0$ and a fixed Abelian group G, $B_m = o$ for $m<0$ (hence $h^m(A)=0$ for all $m<0$).

It is to be emphasized that the above homotopical definition of cohomology (without or with dimension axiom) and the whole algebraic machinery including the spectral sequence is entirely independent of any type of cellular structure of the spaces involved or of "cohomology" in any of the classical meanings.

6. Polyhedra, simplicial cohomology

The special role of polyhedral spaces in the context of cohomology lies, roughly speaking, in the particular "possibilities for effective computation" given by the cellular structure. The precise meaning of this feature, in addition to the general algebraization outlined above, can be described as follows.

By a "polyhedron" we mean here, for simplicity, a finite simplicial complex (the finiteness is only used for part of the following considerations,

[1] The notations normally used are h^m for our h_+^m and $\bar{h}^m$ for our h^m.

and moreover everything can be carried over to spaces having the homotopy type of CW-complexes). For such a polyhedron A, let $A^{(m)}$ denote its m-skeleton, $A^{(-1)}=o\subset A^{(0)}$, and α_m the embedding $A^{(m-1)}\to A^{(m)}$ for $m=0,1,\ldots,n(=\dim.A)$. Then the map $\alpha:o\to A$ is factored as $\alpha=\alpha_n\alpha_{n-1}\ldots\alpha_1\alpha_0$. The spectral sequence for $h^*(\alpha)=h^*(A)$ corresponding to that factorization begins with $E_1=\sum_{m=0}^{n}h^*(\alpha_m)$. By the excision property for cofibrations we have $h^q(\alpha_m)=h^q(A^{(m)}/A^{(m-1)})$; now $A^{(m)}/A^{(m-1)}$, the m-skeleton where the $(m-1)$-skeleton is identified to a single point o, is the union of a certain number of spheres S_m with one common point o, each sphere corresponding to an m-simplex σ_m of A. The elements of the group $h^q(\alpha_m)\cong\Pi(A^{(m)}/A^{(m-1)},B_q)$ can thus be considered as functions associating to each simplex σ_m of A an element of $\Pi(S_m,B_q)=\pi_m(B_q)=h^q(S_m)\cong h^{q-m}(S_0)$; in other words, $h^q(\alpha_m)$ is the group $C^m(A;\pi_m(B_q))=C^m(A;h^{q-m}(S_0))$ *of all simplicial m-cochains of A with values in* $\pi_m(B_q)\cong h^{q-m}(S_0)$. Thus the spectral sequence begins with

$$E_1=C^*(A;h^*(S_0));$$

more precisely, writing p for $q-m$, $E_1=\sum_{m=0}^{n}\sum_{p=-\infty}^{+\infty}E_1^{m,p}$ with $E_1^{m,p}=h^{m+p}(\alpha_m)=C^m(A;\ h^p(S_0))$ [contributing through the spectral sequence to $h^{m+p}(\alpha)=h^{m+p}(A)$]. The differential d_1 in E_1 is easily computed: it is the ordinary simplicial coboundary in $C^m(A;h^p(S_0))$. Therefore we have

$$E_2=H^*(A;h^*(S_0)),$$

more precisely $E_2=\sum_{m,p}E_2^{m,p}$ with $E_2^{m,p}=H^m(A;h^p(S_0))$. Here H^m(and H^*) denote the (reduced) *elementary simplicial cohomology* of the polyhedron A, with the respective coefficient groups, computed as usual from incidence numbers, cocycles and coboundaries.

We summarize the result: *For a cohomology theory* h^* (with respect to a given spectrum $\mathfrak{B}$) *and a polyhedron A there is a spectral sequence for* $h^*(A)$ *with* $E_2=H^*(A;h^*(S_0))$, *where all* $H^m(A;h^p(S_0))$ *with* $m+p=q$ *contribute to* $h^q(A)$. *Thus the first step of the computation of* $h^*(A)$ *is given by the simplicial cohomology of A with coefficients in* $h^*(S_0)$.

The remainder of this lecture (sections 7 and 8) is devoted to a rapid description of some of the numerous applications of this fundamental fact of polyhedral algebraic topology; I have to choose from a large variety of results which increases every day. Some of them use further "partly computable" spectral sequences obtained from other factorizations of maps (the factorization being defined with the help of skeletons), e.g. for fiber maps. A different approach to such spectral sequences, for abstractly given cohomology is due to Dold (an outline can be found in [3]); it includes multiplicative structures and interesting general properties of cohomology for polyhedra.

7. Special cases and remarks

(1) We first consider, for a polyhedron A, the case of a cohomology theory h^* *with dimension axiom*: $h^q(S_0)=0$ for $q\neq 0$, and we write G for the group $h^0(S_0)\cong h^m(S_m)$ for arbitrary $m\geqslant 0$ (the isomorphism being given by suspension). Then in the spectral sequence for $h^*(A)$ we have $E_2=\sum_{m=0}^{n}E_2^{m,p}$ $(n=\dim.A)$ with $E_2^{m,p}=H^m(A;h^p(S_0))=0$ for $p\neq 0$ and $=H^m(A;G)$ for $p=0$ It follows that $d_r=0$ for all $r\geqslant 2$, hence

$$E_2^{m,p} = E_r^{m,p} = E_\infty^{m,p} = 0 \text{ for } p \neq 0, \; = H^m(A; G) \quad \text{for} \quad p = 0, \; m = 0, \ldots, n,$$

and $h^q(A) \cong E_\infty^{q,0}$; we thus obtain

$$h^q(A) \cong H^q(A; G), \quad \text{with} \quad G = h^0(S_0),$$

for $0 \leqslant q \leqslant n$ and 0 otherwise. The isomorphism (of degree 0) $h^*(A) \cong H^*(A; h^0(S_0))$ is the *uniqueness theorem for cohomology with dimension axiom* in the category of polyhedra: any such theory is isomorphic to the simplicial cohomology relative to the coefficient group $G = h^0(S_0)$. It may also be called the simplicial computation theorem for cohomology with dimension axiom, thus making the whole classical machinery available for applications where the explicit structure of cohomology groups is needed.

In terms of the homotopical definitions the dimension axiom tells that the spectrum $\mathfrak{B}$ of the cohomology theory h^* consists of Eilenberg–MacLane spaces $K(G,m)$, i.e., $h^m(A) = \Pi(A, K(G,m))$. Then the above isomorphism $\Pi(A, K(G,m)) \cong H^m(A; G)$ yields the "simplicial computation of homotopical Eilenberg-MacLane cohomology". It should be added that in the whole argument above the spectral sequence is not really needed, since the term E_2 only occurs; the triple sequence for two maps $A^{(m-1)} \to A^{(m)} \to A^{(m+1)}$ would be sufficient. A close inspection of course provides an explicit homotopical description not only of simplicial cohomology classes, but also of simplicial cocycles and coboundaries (useful, for example, in obstruction theory).

In the following H^* always denotes cohomology with dimension axiom, which on polyhedra is identified with simplicial cohomology.

(2) The spectral sequence for $h^*(A)$, where h^* is a general cohomology theory and A a polyhedron, depends functorially upon A and upon h^*. As an application of the functorial dependence upon A, we consider a map $\alpha: A \to A'$ and assume that it induces an isomorphism $H^*(A'; G) \cong H^*(A; G)$ for any G (or, equivalently, that α induces an isomorphism of simplicial integral homology). Then α induces isomorphisms of all the terms E_r, $r \geqslant 2$, of the spectral sequence for $h^*(A')$ onto those for $h^*(A)$, and thus an isomorphism $h^*(A') \cong h^*(A)$. [An alternative proof for this result, not using the spectral sequence, is given by the fact that the map α yields a homotopy equivalence of ΣA with $\Sigma A'$.] The functorial behavior with respect to h^* appears in the next item.

(3) Let $\mathfrak{B} = (B_m, w_m)$ and $\mathfrak{B}' = (B'_m, w'_m)$ be two spectra, and $t: \mathfrak{B} \to \mathfrak{B}'$ a map of $\mathfrak{B}$ into $\mathfrak{B}'$, i.e., a sequence of maps $t_m: B_m \to B'_m$ compatible with the homotopy equivalences w_m and w'_m. Let h^*, k^* denote the cohomology theories relative to $\mathfrak{B}$ and $\mathfrak{B}'$ respectively. For an arbitrary space A, the maps T_A of $h^m(A) = \Pi(A, B_m)$ into $k^m(A) = \Pi(A, B'_m)$ induced by the t_m are then homomorphisms. They constitute a system of natural transformations of the functors h^m into k^m for all m, compatible with the "connecting isomorphism" (cf. section 4). Such a system is called in short a *transformation of the cohomology theory h^* into k^**; any such transformation can be given by maps of the corresponding spectra. As particular cases, these transformations include (additive) cohomology operations, coefficient relations in ordinary cohomology, etc. The spectral sequences for h^* and k^* can be applied to such transformations (there is a possibility of applying also the dual exact and spectral sequences of the $\Pi_n(A, B)$ to $t_m: B_m \to B'_m$, but we do not describe it here).

We first consider a transformation $T: h^* \to k^*$ of cohomology theories which is an *isomorphism on* S_0. *Then, for any polyhedron* A, T induces isomorphisms of the terms $E_r, r \geqslant 2$, of the spectral sequence for $h^*(A)$ into those for $k^*(A)$, and thus T *is an isomorphism* $h^*(A) \cong k^*(A)$. Again it should be observed here that the result can also be obtained by homotopical arguments not using the spectral sequence; namely, it follows easily from the assumption that the map $t: \mathfrak{B} \to \mathfrak{B}'$ corresponding to the transformation is for each m a weak homotopy equivalence $B_m \to B'_m$.

(4) The following procedure due to Dold [3] yields an interesting transformation T relating an arbitrary cohomology theory h^* to ordinary cohomology H^* relative to the coefficient group $h^*(S_0) \otimes \mathbf{Q}$ (tensor product with the group of rationals).

Let k^* be a cohomology theory, with spectrum $\mathfrak{B} = (B_m, w_m)$, and consider the composition of maps

$$\Sigma^N A \xrightarrow{f} S_{N+p} \xrightarrow{g} B_{N+p+q}.$$

For N sufficiently large the homotopy class of f is (under the suspension isomorphism) independent of N and defines an element of the stable cohomotopy group $\pi^p_{st}(A)$, while g defines, for an arbitrary N, an element of $k^q(S_0)$ and gf an element of $k^{p+q}(A)$. We thus get a transformation T' given by

$$T'_A: \pi^p_{st}(A) \otimes k^q(S_0) \to k^{p+q}(A).$$

Now $\pi^p_{st}(S_0)$ is finite for $p \neq 0$ and infinite cyclic for $p = 0$; therefore, if we assume $k^*(S_0)$ to be a module over the ring $\mathbf{Q}$ of rationals, T'_{S_0} will be an isomorphism and thus T'_A an isomorphism for polyhedra A. This result can be applied (a) to $k^q = h^q \otimes \mathbf{Q}$, where h^* is an arbitrary cohomology theory, and (*b*) to $k^q = \sum_{m+p=q} H^m(A; h^p(S_0) \otimes \mathbf{Q})$, and thus yields an isomorphism

$$T''_A: h^q(A) \otimes \mathbf{Q} \cong \sum_{m+p=q} H^m(A; h^p(S_0) \otimes \mathbf{Q})$$

for polyhedra A. Combining this with the natural embedding $\tau_A: h^q(A) \to h^q(A) \otimes \mathbf{Q}$ *we get a transformation* T *given by*

$$T_A: h^q(A) \to \sum_{m+p=q} H^m(A; h^p(S_0) \otimes \mathbf{Q})$$

which reduces to τ_{S_0} *on the* 0-*sphere and which is, for any polyhedron* A, *an isomorphism if tensored by* $\mathbf{Q}$. Thus for polyhedra the difference between H^* and an arbitrary cohomology theory h^* lies essentially only in torsion elements. The transformation T is, by the way, uniquely determined by the fact that T_{S_0} is the embedding $\tau_{S_0}: h^*(S_0) \to h^*(S_0) \otimes \mathbf{Q}$.

(5) There is a famous example (Grothendieck-Atiyah-Hirzebruch) of a cohomology theory without dimension axiom and the corresponding transformation into ordinary rational cohomology. It is given by the *unitary spectrum* $\mathfrak{B} = (B_m, w_m)$ with $B_m = \Omega U$ if m is even, $B_m = U$ if m is odd, where U denotes the infinite unitary group and ΩU its loop-space; by the Bott periodicity theorem ΩU can be used as classifying space for U (there is a (weak) homotopy equivalence $w: U \to \Omega\Omega U$; hence we can put $w_m = 1$ for m

even, $w_m = w$ for m odd). The corresponding cohomology groups $\Pi(A, \Omega U)$ and $\Pi(A, U)$ are denoted by $K^m(A)$ (reduced, usually written $\tilde{K}^m$); one has $K^m(S_0) = \pi_0(\Omega U)$ or $\pi_0(U)$ respectively, i.e. $K^m(S_0) = \mathbf{Z}$ if m is even, $=0$ if m is odd. The transformation T of (4) yields isomorphisms for polyhedra A

$$\begin{aligned} K^q(A) \otimes \mathbf{Q} &\cong \sum_{m+p=q} H^m(A; K^p(S_0) \otimes \mathbf{Q}) \\ &= \sum_{m \text{ even}} H^m(A; \mathbf{Q}) \quad \text{if} \quad q \text{ is even}, \\ &= \sum_{m \text{ odd}} H^m(A; \mathbf{Q}) \quad \text{if} \quad q \text{ is odd}, \end{aligned}$$

thus expressing $K^q(A) \otimes \mathbf{Q}$ by the Betti numbers of A. Moreover one has, of course, the whole spectral sequence of section 6 for $K^*(A)$, with $E_2 = H^*(A; K^*(S_0))$, as given originally by Atiyah-Hirzebruch. All this and related further facts have been used successfully to compute K^* and K^*_+ (unreduced theory, cf. section 5) for complex and real projective spaces, and "truncated" ones; and also to compute, in these cases, cohomology operations $K^* \to K^*$ defined by linear representations of the unitary groups. On the basis of these computations, Adams has given the complete solution of the real and of the unitary vector field problem on spheres (determination of the maximum number of independent fields; for the real case, see [1]).

We have not so far mentioned the relationship between $K^*(A)$ and unitary *vector bundles* over A: Since B_m is the classifying space of the unitary group (m even) or its loop-space (m odd), the group $K^m(A) = \Pi(A, B_m) = \Pi(\Sigma A, B_{m+1})$ is the group of stable unitary vector bundles over A (m even) or over ΣA(m odd); and $K^m_+(A)$ the group of vector bundles with specified fiber dimension. The transformation $T: K^*(A) \to H^*(A; K^*(S_0) \otimes \mathbf{Q})$ above, regarded as a transformation into $H^*(A; \mathbf{Q})$ is nothing else than the "Chern character": for $\xi \in K^*(A)$, $T(\xi) = \text{ch } \xi$ is the Chern character belonging to the total Chern class of ξ (ch ξ and the total Chern class determine each other).

This relation to bundle theory is irrelevant for the computations above, but of course it plays an important part in many other connections; in particular when multiplicative structures are involved, see section 8.

(8) We close this section with some remarks on *duality*. The duality aspects of the two-space homotopy groups $\Pi_n(A, B)$, including their algebraic properties as described in sections 2–4, are obtained by interchanging A and B, i.e, by reversing the direction of all maps. The Hurewicz homotopy groups $\pi_m(B) = \Pi(S_m, B)$ and the cohomology groups $h^m(A) = \Pi(A, B_m)$ for some spectrum $\mathfrak{B} = (B_m, w_m)$ behave, in some respects, dually to each other: from $\Pi(A, B)$ one obtains the π_m by choosing $A = S_m$, the h^m by choosing $B = B_m$ (with equivalences $w'_m: \Sigma S_m \to S_{m+1}$ and $w_m: B_m \to \Omega B_{m+1}$, hence $\pi_m(B) = \Pi_n(S_{m-n}, B)$ and $h^m = \Pi_n(A, B_{m+n})$, independent of n), and thus many dual features obviously are carried over from the general $\Pi_n(A, B)$ to a dual behavior of the π_m and the h^m (exactness, spectral sequence for composition of maps, excision, etc.). It is, however, clear that cohomology goes beyond the general set-up and that there is no obvious dualization of the special properties of h^* described in sections 5 and 6. In this context we add two simple comments.

(a) The simplicial computation theorem for h^* relies on the special struc-

ture of polyhedral spaces and on the excision property for cofibrations. In order to obtain a similar computation scheme for π_* one has to take, instead of polyhedra, complicated spaces constructed artificially (by certain fibrations with Eilenberg–MacLane fibers). There is no evidence so far that these spaces are of geometrical significance.

(*b*) If one considers, in a spectrum $\mathfrak{B}=(B_m, w_m)$ a fixed B_m, one has in $\mathfrak{B}$ a "trivial" part below B_m, in the sense that all B_n with $n<m$ can be replaced by $\Omega^{m-n}B_m$, and a non-trivial part above B_n. The system of spheres S_m together with $w_m': \Sigma S_m = S_{m+1}$ contains for $m=0$ *only the trivial part above* S_0, $S_n=\Sigma^n S_0$ for $n>0$, and it cannot be extended to a non-trivial part below S_0. Thus the π_m do not go beyond the general $\Pi_n(A,B)$, except that a special space A is being used; other choices of A such as Moore spaces (π_m with coefficients, cf. [4]) do not essentially change the situation. However a more systematic use of spectra may lead to further duality.

It is well-known that the π_m and the $\Pi_n(A,B)$ in general, even if there is no simple geometrical class of spaces for which they are easily computable, have been and will continue to be of great interest. This is due both to their intuitive geometrical meaning and to their algebraic structure as explained above, which through the spectra directly leads to cohomology. There are various ways to bring also *homology* and related duality theorems into the picture but we do not describe them here (cf. G. W. Whitehead [8]).

8. Multiplicative cohomology

In this section we sketch some aspects of general cohomology which play a central role in current developments of algebraic topology and which, at least in their algebraic background, belong to the methods outlined in this lecture. No details will be given, and very few references and definitions will be included.

A cohomology theory is *multiplicative* if h^* is provided with a natural product turning this graded group into a graded anti-commutative ring, with $h^*(A^+)=h^*_+(A)$ having a unit $e\in h^0(A^+)$, and if, moreover, the relative cohomology $h^*(\alpha)$, $\alpha: A'\to A$, is a module over $h^*(A)$. (In particular, if P denotes a one-point space, $h^*(P^+)=h^*(S_0)$ has a unit $e\in h^0(S_0)$; under suspension it corresponds to an element γ^n of $h^n(S_n)$.)

The spectral sequence arguments for polyhedra can be carried over to multiplicative cohomology theories. In particular, a transformation T: $h^*\to k^*$ where h^* and k^* are multiplicative cohomology theories and k^* a $\mathbf{Q}$-module, preserves products if it does so on S_0. For example, $H^*(\ ;\mathbf{Q})$ is multiplicative with respect to the cup-product; and K^* with respect to the product given in K^0 by the tensor product of bundles and given between any K^m and K^n by means of standard suspension properties. Then the Chern character ch: $K^*\to H^*(\ ;\mathbf{Q})$ is a ring homomorphism on S_0, namely the embedding of $K^*(S_0)$ into $K^*(S_0)\otimes\mathbf{Q}$, and ch is a product preserving transformation on polyhedra.

As one very well knows, further very effective structure properties of ordinary cohomology H^* (with a coefficient ring, for example) are available if one confines attention to *manifolds*; most important are of course those

connected with Poincaré duality and the Hopf "inverse homomorphism" $f_*: H_+^*(A) \to H_+^*(A')$ associated with a map $f: A \to A'$ of manifolds. In particular, if f is the trivial map $A \to P$ of the n-dimensional orientable manifold A onto the one-point space P, then $f_* a = [a] \cdot e \in H_+^0(P)$, where $[a]$ is the "top-component" $\in H^n(A)$ of a, evaluated on the manifold-cycle of A. Such special properties valid for manifolds can be expressed with the help of the tangent bundle or normal bundle of A (for differentiable manifolds; for non-differentiable ones the new concept of "micro-bundle" has to be used, cf. the lecture of John Milnor). In order to obtain similar further structure properties for general cohomology h^*, it is therefore natural to start from *real vector bundles over A.*

Let ξ be such a bundle, with fiber dimension n, and h^* a multiplicative cohomology theory. Let further A^ξ denote the Thom space of ξ (the associated solid sphere bundle E over A modulo its boundary S_{n-1}-bundle $\dot{E}$), p the projection $E \to A$. Using the multiplicative properties of the spectral sequence (cf. Dold [3]) one proves: *If* there is a $v \in h^n(A^\xi)$ which on a fiber of A^ξ reduces to an element corresponding to $\gamma^n \in h^n(S_n)$, then the map $\varphi_h: h_+^m(A) \to h^{m+n}(A^\xi)$ given by $\varphi_h(a) = p^*(a) \cdot v$, $a \in h_+^m(A)$ is an isomorphism (Thom–Gysin isomorphism). The element v, and therefore also φ_h, are not unique in general. [For ordinary cohomology $H^*(\ \ ; \mathbf{Z})$ the element v exists (and is unique) if and only if ξ is orientable (and oriented); Poincaré duality can be obtained as a special case. For K^* special conditions for existence and uniqueness of v have been given by Atiyah–Hirzebruch.] If one now takes for A a (differentiable) manifold and for ξ the stable normal bundle ν of A, the Thom–Gysin isomorphism $\varphi_h: h_+^*(A) \to h^*(A^\nu)$ can be used to define an inverse homomorphism: A^ν is a model of the S-dual $\hat{A}$ of A, and if $f: A \to A'$ is a map of differentiable manifolds, there is an S-dual map $\hat{f}: \hat{A}' \to \hat{A}$ and an induced homomorphism $\hat{f}^*: h^*(\hat{A}) \to h^*(\hat{A}')$; the inverse homomorphism f_* is then defined by $f_* = \varphi_h'^{-1} \hat{f}^* \varphi_h$: $h_+^*(A) \to h^*(\hat{A}) \to h^*(\hat{A}') \to h_+^*(A')$. Here φ_h' of course denotes the Thom–Gysin isomorphism $h^*(A') \to h^*(A'^{\nu})$; note that $f_*: h_+^*(A) \to h_+^*(A')$ depends on the choice of $v \in h^*(A^\nu)$ and $v' \in h^*(A'^{\nu})$. This inverse homomorphism in h^*-theory has properties analogous to those of the Hopf inverse homomorphism for H^*. The element $f_* a \in h_+^*(A')$ can be considered as the generalized analogue of the "top-component" of a, in particular for $A' = P$.

Using the Thom–Gysin isomorphism $\varphi_h: h_+^*(A) \to h^*(A^\nu)$, if it exists, one further defines, for a product-preserving transformation $T: h^* \to k^*$ of multiplicative cohomology theories, a sort of "genus" $\mathfrak{T}(A)$ of the manifold A: one puts

$$T\varphi_h(e) = \varphi_k(\mathfrak{T}(A)),$$

where $e \in h_+^*(A), \mathfrak{T}(A) \in k_+^*(A)$. It then easily follows that

$$T\varphi_h(a) = \varphi_k(T(a) \cdot \mathfrak{T}(A))$$

for all $a \in h_+^*(A)$. Combining this with the inverse homomorphisms $f_*^h: h_+^*(A) \to h_+^*(A')$ and $f_*^k: k_+^*(A) \to k_+^*(A')$ for a map of manifolds $f: A \to A'$, a simple computation yields the formula

$$f_*^k(T(a) \cdot \mathfrak{T}(A)) = T(f_*^h a) \cdot \mathfrak{T}(A'),$$

i.e., a relation for the "top-components" $f_*^k(Ta)$ and $T(f_*^h a)$, $\mathfrak{T}(A)$ and $\mathfrak{T}(A')$ giving the necessary corrections between the two terms. [Of course $\mathfrak{T}(A)$ and $\mathfrak{T}(A')$ depend on the same choices of the Thom–Gysin isomorphisms φ_h, φ_k on A and A' as the f_*^h and f_*^k]. This general relation is a formal expression for the various Riemann–Roch theorems, and also for the analogous formulae for cohomology operations, according to the choices of h^*, k^* and T, and to the special properties of f. It was communicated to the author by E. Dyer, but seems already to belong to the folklore of algebraic topology. In the different interesting cases to which this general formula can be applied, deeper results are based on explicit relations of $\mathfrak{T}(A)$ to characteristic classes.

REFERENCES

[1]. Adams, J. F., Vector fields on spheres. *Ann. Math.*, 75 (1962), 603–632.

[2]. Brown, E. H., Cohomology theories. *Ann. Math.*, 75 (1962).

[3]. Dold, A., Relations between ordinary and extraordinary homology. *Algebraic Topology Colloquium Aarhus*, 1962, 2–9.

[4]. Eckmann, B. & Hilton, P. J., Groupes d'homotopie et dualité. *C. R. Acad. Sci. Paris*, 246 (1958), 2444–2447; 2555–2558; and 2991–2993. See also: Eckmann, B., Groupes d'homotopie et dualité. *Bull. Soc. Math. France*, 86 (1958), 271–281.

[5]. Eckmann, B. & Hilton, P. J., Group-like structures in general categories. I. Multiplications and comultiplications. *Math. Ann.* 145 (1962), 227–255.

[6]. Eckmann, B. & Huber, P. J., Spectral sequences for homology and homotopy groups of maps. *Report Battelle Memorial Institute Geneva*, (1960).

[7]. Huber, P. J., Homotopy theory in general categories. *Math. Ann.*, 144 (1961), 361–385.

[8]. Whitehead, G. W., Generalized homology theories. *Trans. Amer. Math. Soc.*, 102 (1962), 227–283.

AUTOMORPHIC FUNCTIONS AND THE THEORY OF REPRESENTATIONS

By I. M. GEL'FAND(1)

1. Introduction

According to F. Klein the investigation of geometrical objects may be reduced to the study of properties invariant under a group of transformations, i.e. to the study of homogeneous spaces.

By a homogeneous space X we mean any manifold acted upon by a group of transformations which is a Lie group. We can therefore study in a unified manner the whole class of the most symmetrical objects, aesthetically perfect, such as sphere, Lobachevsky space, Grassmann's manifold, the space of positively-definite matrices, etc.

It is rather remarkable that we can reach the same aesthetic perfection in the study of the set of functions $f(x)$ defined on the homogeneous space X. A transformation $x \to xg$ in X gives rise to a linear operator $T_g f(x) = f(xg)$ in the function space. So we are led to a representation of the group G, since the product of the transformations T_{g_1} and T_{g_2} corresponds to the product of the elements g_1 and g_2 of G.

Roughly speaking the problem is to decompose the function space into minimal invariant subspaces, or, which is the same, to decompose the representation into irreducible representations and to study invariant families involved.

The compact case of this problem (i.e. compact X and G) was investigated by H. Weyl and E. Cartan who took as model the rotation group. The invariant families for the rotation group consist of spherical functions. The least invariant systems of functions, arising in the general case, we shall also call the spherical functions on X. But only after one rejects the compactness condition and passes over to infinite-dimensional representations can one fully appreciate the importance and interest of this problem.

In this report I should like to tell about some results obtained by me in collaboration with my friends I. I. Pyatezki-Shapiro and M. I. Graev. Some of the results were inspired by important works of Selberg and Godement. It should be said that only through systematic employment of the theory of infinite-dimensional representations can one obtain complete understanding of these results.

Each homogeneous space X is associated with a group G and its subgroup Γ (stationary subgroup) consisting of those elements of G which leave immobile some fixed point $x_0 \in X$. The first part of the report is devoted mainly to the consideration of the case in which this subgroup Γ is a discrete group and the space X has finite volume. The functions relating to this case we shall call automorphic functions. Thus automorphic functions are spheri-

(1) The author dedicates this report to his friends, L. Schwartz, M. Morse, J. Leray, O. A. Oleinik, J. G. Petrovskiy, L. Bers, R. Courant, P. Lax, whose kind assistance has greatly helped him.

cal functions, associated with a discrete stationary subgroup. Automorphic functions in this sense were introduced in 1950 in a paper by S. V. Fomin and myself [3]. The ordinary automorphic functions are included among these if we take the group of real 2×2-matrices.

Permit me to cite another interesting and important example. Let us be given an arbitrary Riemann surface with a Riemann metric on it. Consider the manifold X the points of which are taken to be linear elements (i.e. points with directions) of the initial surface. Suppose that the universal covering space of our Riemann surface is the upper halfplane. Then the group G acting on X will be the group of all linear fractional transformations with real coefficients and the stationary subgroup Γ will be one of its discrete subgroups. If the original Riemann manifold is compact, i.e. if it is associated with an algebraic function, then the resulting representation of G in $L_2(X)$ can be decomposed into a countable direct sum of representations. Every irreducible representation of the group G under consideration can be characterised by a set of indices which we shall call the number of the representation. The numbers of the representations involved in the decomposition are invariants of the Riemann surface. They appear to form a complete system of invariants, and so it may happen that this question is connected with the classical problem of moduli in which such remarkable success was attained recently by Bers and Ahlfors. We can prove a weaker result: every continuous deformation of Riemann surface, leaving unchanged these invariants, is the identity.

So in the first half of the report we consider some general questions about the decomposition of a representation in a function space $L_2(X)$ on a homogeneous space X with discrete stationary subgroup. The characterization of the resulting representations and the method of decomposition lead us to very interesting functions which we call Zeta-functions of the given homogeneous space. This system of functions is closely connected with the classical Zeta-function. In the case of the group of real 2×2-matrices we can reduce these functions, using the duality theorems, to the functions, which were introduced by Selberg from particular considerations.

Now we wish to say a word about the methods. M. I. Graev and I have proposed [1] the so-called method of horyspheres for the investigation of representations in $L_2(X)$ in the case in which Γ is continuous, but not discrete as in the case discussed above. This method can be applied with suitable modifications also to the case considered here. This method leads us to a series of theorems concerning the structure of the representations and to the Zeta-function of a homogeneous space. But the general idea of the horysphere method is not sufficient. It is necessary to study the operators $\int F(g)T_g dg$ in detail, where $F(g)$ is a finite function. This part of our work reminds one strongly of the theory of the S-matrix in quantum mechanics. The analogy is intrinsic. For example the Zeta-functions of a homogeneous space are quite analogous to the Heisenberg S-matrix.

In the second part of the report we consider analogous questions concerning the structure of representations in $X = G/\Gamma$ in the case of a semisimple Lie group over a finite field (in Chevalley-Dickson sense). So it is necessary to describe the representations of Lie groups over finite fields. These questions could serve as the theme of a separate report and we are forced to give only a short account of them. We obtain here a interesting system of Zeta-functions for a homogeneous space above a finite field.

The author has chosen as the theme of his report developments that are in the very earliest stages.

We hope that the many interesting problems and relationships which arise will compensate for the unavoidable lack of a description clear and comprehensible in every detail.

We hope that a reasonable bourbakisation of all facts given in this report and of other facts and problems of the theory of representations will lead to the creation of a domain of algebraic functional analysis in which these facts will be the main examples.

2. The case of a compact homogeneous space

If G is a semisimple Lie group, Γ a discrete subgroup and X the space G/Γ of left cosets, then evidently to each element $g \in G$ there corresponds a movement in X transforming x to xg. Denote by $L_2(X)$ the set of all functions on X with integrable square. To each element $g \in G$ we make correspond the unitary shift operator T_g in $L_2(X): T_g f(x) = f(xg)$. The operators T_g form a representation of the group G.

The main problem is to decompose this representation into irreducible ones.

The operators $T_\varphi = \int_G \varphi(g) T_g dg$ where $\varphi(g)$ is a finite function on G play an important role in the theory of infinite-dimensional representations. In our situation the use of operators T_φ is based upon the fact that for every continuous finite function $\varphi(g)$ the operator T_φ is an integral operator in X with the kernel

$$K(x_1, x_2) = \sum_{\gamma \in \Gamma} \varphi(g_1^{-1} \gamma g_2), \tag{1}$$

where g_1 and g_2 are representatives of cosets x_1 and $x_2 \in X$. It is easy to verify that the series for the function $K(x_1, x_2)$ converges uniformly in every compact domain. Thus if X is compact the operator T_φ is completely continuous for every finite continuous function $\varphi(g)$.

It is easy to prove the following general proposition.

If the unitary representation $g \to T_g$ of Lie group G in a space H is such that operator T_φ is completely continuous for every finite continuous function $\varphi(g)$, then H can be decomposed into a countable sum of irreducible representations (unitary) of the group G, where the multiplicity of each irreducible representation is finite.

It follows at once from this proposition that when X is compact the representation in $L_2(X)$ is decomposable into a countable sum of irreducible unitary representations of the group G. This fact does not exhaust all information about the irreducible representations contained in $L_2(X)$ which can be obtained by the use of the operators T_φ. There is a formula which gives in a reasonable sense a complete description of all irreducible unitary representations contained in $L_2(X)$.

Let $H_1, H_2, \ldots$ be irreducible non-equivalent representations contained in $L_2(X)$; $N_1, N_2, \ldots$, their multiplicities. Let $\pi_k(g)$ be the character of the irreducible representation H_k. The existence of such characters was proved by Godement and Harish Chandra. Then the following important formula

$$\sum_{m=1}^{\infty} \int_{F_m} \varphi(g^{-1}\gamma_m g)\, dg = \sum_{k=1}^{\infty} N_k \int \varphi(g)\, \pi_k(g)\, dg \tag{2}$$

is valid, where $\gamma_1, \gamma_2, \ldots$ is a sequence of mutually non conjugate elements of the group Γ and F_m is a fundamental domain for the centralizer Γ_m of the element γ_m. This formula makes it often feasible to find out what representations enter in $L_2(X)$, and with what multiplicities. A particular case of this formula (namely the case in which the functions $\varphi(g)$ are invariant under U—the maximal compact subgroup of the semisimple group) can be essentially reduced by the use of the so-called duality theorem [2] to the previously known formula of Selberg [4].

Formula (2) may be used to investigate the asymptotic distribution of "numbers" of the representations, occurring in $L_2(X)$. The leading term of the asymptotic expansion will be obtained if we choose the function $\varphi(g)$ to be concentrated in a decreasing sequence of neighbourhoods of the unit of the group. Now we give following example.

Consider the asymptotic behaviour not of all representations but only of the so-called representations of class I. By definition these representations are the representations entering in the decomposition of $L_2(G/U)$ where U is a maximal compact subgroup of G.

These representations are given in the following manner. Let $\mathfrak{A}$ be Cartan subalgebra of the symmetric space G/U and let $\mathfrak{A}^+$ be the cone of the positive vectors in $\mathfrak{A}$ (i.e. vectors for which $(\alpha, \alpha) \geqslant 0$ for all positive roots α). Then every representation of class I is uniquely determined by some vector in $\mathfrak{A}^+$. If now X is a compact homogeneous space G/Γ then the representations of class I entering into $L_2(X)$ form a countable set of points in $\mathfrak{A}^+$. Their asymptotic distribution is given by the following formula. Denote by α the positive roots of the symmetric space G/U and by v_α the multiplicities of the roots of G/U. Then the formula

$$N(B_n) \sim C_G C_\Gamma \int_{B_n} \prod_{\alpha>0} (\varrho, \alpha)^{v_\alpha}\, d\varrho \tag{3}$$

is valid, where B_n runs through an increasing sequence of subregions in $\mathfrak{A}^+$; C_G is a constant depending only on G; C_Γ is the volume of space $X = G/\Gamma$ and $N(B_n)$ is the number of irreducible representations entering in $L_2(X)$ indices of which belong to B_n. The proof is based on some results of F. I. Karpelevich and S. G. Gindikin.

Apparently a similar formula is valid for the other types of irreducible representations.

Since for the group of 2×2-matrices as was said, the "numbers" of the representations are invariants of the Riemann surface, the terms of the asymptotic expansion are also such. It would be very interesting if a few leading terms of this expansion were to play the role of moduli. The next terms of asymptotics demand apparently more averaging than the mere evaluation of $N(B_n)$.

Classical automorphic forms enter here in natural way. Among the representations of the group G there are isolated ones, i.e. those occurring in isolated form already in $L_2(G)$. Among the isolated ones there are representations which are realizable by analytic functions [11, 12, 13]. These isolated representations, if they do occur in the decomposition of $L_2(X)$, are con-

nected with automorphic forms in natural way. It is interesting to observe that there are still other isolated representations (Graev [6]). For every isolated representation one can deduce from the spur formula an explicit formula for the multiplicity N_k. In the case of classical automorphic forms such a formula was obtained by Hirzebruch [14] who used the Riemann–Roch theorem and by Selberg [4] who used a method close to ours.

3. Horyspheres and regular subgroups

The method of horyspheres, which has been elaborated elsewhere, mainly in the paper [1], is effective when X is not compact. In general the method is as follows. Let X be a homogeneous space acted upon by a group G. With the space X one can associate a space Ω whose elements are taken to be surfaces in X which we shall call horyspheres (the definition of horysphere will be given later). To each function defined on X we make correspond its integrals over horyspheres, then the representation in the space of functions on X maps homomorphically onto a representation in the space of functions on Ω. As a consequence the problem of decomposing the representation in the space of functions on X is reduced to the following two problems: (1) to find the kernel of the homomorphism; (2) to decompose the obtained set of functions on Ω into irreducible representations. In some important cases, discussed in [1], the kernel is zero and induced representation on Ω can be given a simple description. But in our case the kernel of homomorphism is not zero and the description is not trivial. Nevertheless the method of horyspheres is very productive in this situation.

In particular it makes it possible to separate quite effectively the series of representations occurring continuously in $L_2(X)$ from the series occurring discretely, or in other words to separate the discrete spectrum from the continuous one.

Let X be an arbitrary homogeneous space; then the set of horyspheres in X is not a homogeneous space itself and has not in general a Hausdorff structure. The structure of the space the points of which are transitive components in the space of horyspheres (i.e. the sets of horyspheres which can be carried one into another by the movements of the group) plays a fundamental role in the description of the spectrum of irreducible representations in $L_2(X)$.

Now we proceed to give the exact definitions. Let G be a real semisimple Lie group and $g(t)$ a one-parameter subgroup. The set Z consisting of all $z \in G$ such that

$$\lim_{t\to\infty} g(-t)\,zg(t) = 1$$

is called the horyspherical subgroup associated with the subgroup $g(t)$. For example if G is the group of real 2×2-matrices then every horyspheric subgroup is conjugate to the subgroup of all matrices of the form $\begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix}$.

If G is the group of real matrices of order $\mathbf{n}$, then there exist as many non-conjugate horyspherical subgroups as there are different partitions of the the number $\mathbf{n}:\mathbf{n} = k_1 + \ldots + k_s$ with positive integers k. (Partitions that are distinguished by the order are to be considered as the different ones.)

Now we proceed to define the horyspheres. Let X be a homogeneous space of a group G.

The orbits of horyspherical groups will be called horyspheres in X. The horysphere will be called compact if the set of points of which it consists, is compact.

Now we proceed to describe the class of discrete subgroups Γ which we shall be concerned with. First of all here belong these for which the volume of G/Γ is finite. It is quite possible that all such subgroups satisfy the assumptions formulated below. For the groups of second order this follows from the results of Siegel. It would be very interesting to prove the assertion in the general case.

Now we introduce the following definition. A set $Y \subset X$ will be called a cylindrical set, if it can be split into mutually non-intersecting compact horyspheres which can be carried one into another by movements. Thus in a cylindrical set all the horyspheres are of the form $x_0 Zg$ where Z is some fixed horyspherical subgroup.

Now we formulate the following fundamental definition.

A discrete subgroup Γ of the semisimple group G will be called regular if the factorspace G/Γ has a finite covering by regular bounded cylindrical sets and the intersection of each pair of them is compact (the definition of regular bounded cylindrical sets is given below).

It is easy to verify that the factor-space G/Γ for each regular discrete subgroup Γ of semisimple group G has finite volume.

We shall observe finally that except for the group of real 2×2-matrices all presently known examples of irreducible discrete subgroups of semisimple Lie groups such that the factor space has finite volume are arithmetical groups that are constructed by the well-known construction of Borel.

We shall call linear algebraic group G every group consisting of all complex $\mathbf{n} \times \mathbf{n}$-matrices the elements of which satisfy given polynomial relations. In the following we shall assume that the coefficients of these polynomials are rational numbers. We denote by G_z the set of all matrices in G whose elements are integers and whose determinant is equal to 1. In a similar way G_R is the set of matrices in G whose elements are real numbers. It is easy to see that G_z is a discrete subgroup of the group G_R.

A. Borel and Harish Chandra [8] have proved that the volume of the factor-space G_R/G_z is finite, if G_R is semisimple group.

Apparently it can be proved by their methods that G_z is a discrete regular subgroup of G_R.

Now we proceed to define the regular bounded cylindrical sets. It is not difficult to see that for every cylindrical set Y there exists a horyspherical subgroup Z and a set $S \subset G$, whose image in G/Γ is Y, with the following properties:

(1) For every $g \in S$ and $z \in Z$ there exists $\delta \in \Delta = \Gamma \cap Z$ such that $\delta z g \in S$.

(2) If $g_1, g_2 \in S$ and $g_1 g_2^{-1} \in \Gamma$ then $g_1 g_2^{-1} \in \Delta$.

The first condition means that Y consists of horyspheres. The second that these horyspheres do not intersect.

If the set Y is compact then one can choose S so that it will have also the following properties:

(3) There exists a neighbourhood of unity U_2 such that if $g_1^{-1} \gamma g_2 \in U$, where g_1, $g_2 \in S$, then $\gamma \in \Delta$.

(This condition is more strong than condition 2.)

(4) For every $z \in Z$ there exists a compact neighbourhood of unity U_z such that $g^{-1} z g \in U_z$ for any $g \in S$.

Condition (4) plays in this theory the fundamental role.

We shall agree to call bounded any cylindrical set Y for which there exist sets S and Z with the properties (1)–(4). Bounded cylindrical sets are in general non-compact and this seems to be the geometrical reason for the existence in semisimple groups of discrete subgroups such that the factor-space G/Γ has finite volume and at the same time is non-compact.

Let Y be a bounded cylindrical set. It can be shown that all elements $g \in S$ are representable in the form $g = zat$, where a belongs to a certain accompanying subgroup A of Z,(1) and t belongs to a certain compact set T in G. We shall agree to call a normal subgroup $\tilde{Z}$ of Z allowable, or simply allowable, if the intersection of the Lie algebra of the group $\tilde{Z}$ with every root subspace is either vacuous or contains the root subspace.

A bounded cylindrical set will be called regular if

(5) For each allowable subgroup $\tilde{Z}$ of Z (including the group Z itself) the factor-space $\tilde{Z}/(\tilde{Z} \cap \Gamma)$ is compact.

Thus finally the regular bounded cylindrical sets are characterized by the existence of the sets S, Z, A for which the conditions (1)–(5) hold.

4. The separation of the continuous spectrum of representation from the discrete one

We shall suppose as formerly that we have a homogeneous space $X = G/\Gamma$, where the discrete group Γ is regular and G is a real semisimple group.

By the use of the horysphere method we can decompose the space $L_2(X)$ into a direct orthogonal sum of two spaces. One of them, which we denote $L_2^0(X)$, is decomposable into a discrete direct sum of irreducible representations. The other is decomposable into representations of continuous spectra. However there may be representations which enter into the second space discretely, the so-called representations of the complementary series. They get into the second space because they are, so to speak, the analytical continuation of the continuous spectrum involved. One more justification of this fact is that they hit exactly at the singular points of Zeta-functions corresponding to the given continuous series. It is very much like the complementary discrete spectrum of quantum mechanics that hits exactly at the zeros of the S-matrix.

The decomposition into the direct sum is carried out in the following way. To every function $f(x) \in L_2(X)$ we make correspond its integrals over *compact* horyspheres and we denote by $L_2^0(X)$ the space of functions which have their integral over any compact horysphere equal to zero; $L_2'(X)$ is the orthogonal complement of $L_2^0(X)$.

Then *the space $L_2^0(X)$ can be decomposed into a countable number of irreducible representations*. The proof is based on a study of asymptotic properties of the kernel of the integral operator $\int \varphi(g) T_g dg$ at the regular cylindrical sets.

---

(1) A commutative subgroup A is called accompanying if each element of A is semisimple and if A is generated by oneparametrical subgroups $g_1(t), \ldots, g_n(t)$ to each of it Z is associated.

5. Functions in the space of horyspheres

We divide the set of compact horyspheres into transitive families Ω_i (i.e. sets of horyspheres which can be carried one into another by movements). When Γ is regular there will be only a finite number of such families. The families Ω_i are partially ordered in a natural way, namely $\Omega_i < \Omega_j$ if each horysphere of Ω_i can be split into horyspheres of Ω_j.

Now we take one of the transitive horysphere families Ω_i and make correspond to each function $f(x) \in L_2(X)$ its integrals over horyspheres of Ω_i. We obtain a function $\check{f}(\omega)$ defined on Ω_i. The inner product defined in $L_2(X)$ is carried in a natural way into the set of these functions. Thus in $L_2(\Omega_i)$ there is defined a quadratic form. To this form there corresponds an operator M defined by

$$\int_{\Omega_i} \check{f}\bar{\varphi}\, d\omega = [\check{f},\ M\varphi],$$

where $[\check{f}_1, \check{f}_2]$ is the inner product inherited from $L_2(X)$. This operator plays a fundamental role. It is permutable with the movements in Ω_i. As a consequence this operator, by the decomposition of representation in $L_2(\Omega_i)$ into irreducible ones, will be, in every system of representations equivalent to the given one, a matrix whose order coincides with the number of equivalent representations. In the general case this number is less than or equal to the order of Weyl group. Thus after the decomposition there will arise a matrix of order equal to the order of Weyl group. (For a group of $\mathbf{n} \times \mathbf{n}$-matrices this order is equal to $\mathbf{n}!$). This matrix depending on the "number" of the representation is by definition the Zeta-function of the space X. These functions, associated with the "number" of representation, are closely connected with such functions as the Riemann Zeta-function and its generalizations. It is not excluded that a deeper development of this theory will throw light on the blank spaces in the theory of the classical Zeta-function. The results just formulated give us a chance to describe the representations involved. It could be done in detail but lack of time prevents us from doing so. We merely point out that the spectrum of the representation of the component, associated with any Ω_i, has Lebesgue type and multiplicity 1. The indices of the representations fill up several linear subspaces of Cartan algebra. The description of the allowable subspaces is carried out by induction. The dimension of these spaces is easy to calculate. It is equal to the dimension of a group which is associated with Ω_i and which consists of all the homeomorphisms permutable with the movements of G in Ω_i.

6. Example: the matrix group G and the subgroup Γ of matrices with integer elements

As an example we take the subgroup Γ of integer elements of the real $\mathbf{n} \times \mathbf{n}$-matrix group. Each transitive family of compact horyspheres is defined by a partition $\mathbf{n} = \mathbf{n}_1 + \ldots + \mathbf{n}_k$. The maximal family Ω_1 is defined by the partition $\mathbf{n} = 1 + \ldots + 1$ and consists of orbits of the triangular group. Every irreducible representation of G can be defined by an index which is a

vector in a space of dimension $\mathbf{n}-1$. The set of indices of all the irreducible representations involved in the decomposition of $L_2(X)$, where $X=G/\Gamma$, is called the spectrum of $L_2(X)$. To each Ω_i there corresponds a spectrum consisting of not more than a countable number of linear subspaces. To Ω_1 itself there corresponds a spectrum of multiplicity 1 filling up the whole space. The spectrum of $L_2(X)$ is the union of the subspaces mentioned above and of a countable number of points.

Now we shall exhibit the Zeta-function corresponding to the homogeneous space G/Γ. We shall restrict ourselves to the Zeta-function of highest dimension. The Weyl group in this case is the symmetric group of degree $\mathbf{n}$. Thus the Zeta-function is a matrix of order $\mathbf{n}!$. We shall write down the first row because the rank of the matrix is equal to 1 and all other rows are easily expressed in terms of the first. (Each representation associated with Ω_1 enters in $L_2(X)$ only one time!) So we have: if σ is a permutation of the symmetric group and $x=(x_1,\ldots,x_{\mathbf{n}})$ is the "number" of a representation $(x_1+\ldots+x_{\mathbf{n}}=0)$, then

$$\zeta_\sigma(x)=\prod_{\substack{i>j\\ \sigma_{(i)}<\sigma_{(j)}}} B\left(\frac{1}{2},\frac{x_i-x_j}{2}\right)\frac{\zeta(x_i-x_j+1)}{\zeta(x_i-x_j)}, \tag{5}$$

where $\sigma=(\sigma_1,\ldots,\sigma_{\mathbf{n}})$.

It is not difficult to write this formula in standard root notation. Written in this form it can be generalized.

The functional equation for the Zeta-function which is a consequence of the general theory of representations, is of the form

$$\zeta_{\sigma_1\sigma_2}(x)=\zeta_{\sigma_2}(x)\,\zeta_{\sigma_1}(\sigma_2 x). \tag{6}$$

7. The representations of finite Chevalley groups

The success of the theory of representations depends after all on a happy construction of irreducible representations. In the case of the complex semisimple Lie groups a simple construction of irreducible representations was given (in 1947–1950 by M. A. Naimark and the author [9] (later we shall speak about the construction). However, if we go to other classes of semisimple groups—the real semisimple groups, the matrix groups over finite fields etc.—the construction of Naimark and author does not yield all the representations but only a small part of them.

Now we shall discuss the classification of the irreducible representations of semisimple matrix groups over a finite field. We shall deal with the groups which were considered in the well-known papers by Dickson [15] and Chevalley [16]. Included here for example are the unimodular matrix groups over finite fields, the matrix groups leaving invariant some quadratic form etc.

Already in the last century in one of the first papers on the theory of representations Frobenius found the characters of the unimodular group of 2×2-matrices over a finite field. Later (in 1928) Hecke gave the construction for half of the irreducible representations of this group. The construction for the other groups was absent.

Now we proceed to construct the irreducible representations. In the case of a complex semisimple Lie group G the construction of the principal series of irreducible representations is as follows. We consider the homogeneous space G/Z of the right cosets of G with respect to a maximal nilpotent subgroup Z. We call this space the principal affine space associated with G. The set of homeomorphisms of G/Z permutable with the movements form a commutative group which we shall call the homothety group. A function defined on G/Γ will be called homogeneous if it is multiplied by a constant under a homothety.

The irreducible representations of the principal series are realized in spaces of homogeneous functions on G/Z; the operator of representation is defined as the shift operator. We have already noted that this construction when transferred to the simple Dickson–Chevalley groups over a finite field, will give us only a small part of all irreducible representations. Now we shall explain how to construct *all* irreducible representations. Again we consider a representation in the space of functions on G/Z where Z is a maximal nilpotent subgroup. An operator of the representation will be defined not as mere shift but as a shift operator multiplied by a fixed function depending on a point of the space and on an element of the group

$$T_\chi(g)f(z) = f(zg)\,\alpha(z,g).$$

It is easy to show that the function $\alpha(z,g)$ is defined essentially by a one-dimensional representation $\chi(z)$ of the group Z. For brevity we shall call the $\chi(z)$ simply characters. It can be shown that by the decomposition of these representations we shall obtain all the irreducible representations. This follows from the fact that in every representation of G there exists a vector which is an eigenvector with respect to the element of Z. Now we introduce a partial ordering into the set of characters: let $\chi_1 < \chi_2$ if from $\chi_2(s) = 1$ it follows $\chi_1(s) = 1$ for each $s \in Z$ such that $s = e^{E_\alpha t}$ where α is a root of G; the maximal characters $\chi(s)$ will be called the characters of general position. It turns out that the representations $T^\chi(g)$ corresponding to the characters χ of general position do contain each irreducible representation of G not more than one time. The irreducible representations entering into these $T_\chi(g)$ we shall call the principal representations of G, all others will be called degenerate. The number of all irreducible representations of G is a polynomial of k, where k is the order of field under consideration. It turns out that the number of degenerate representation is a polynomial of lower degree. In this sense we can say that the principal representations are almost all irreducible representations of G. It can be shown that the dimensions of the principal representations of G can be expressed as polynomials in k of degree N, where N is the dimension of group Z, and that the dimensions of the representations of the degenerate series are expressed as polynomial of lower degree. As an illustration we give a formula for the dimensions of the different principal representations of the unimodular matrix group of the $\mathbf{n}$th order. The principal representations are split into several series. Each series is defined by a partition $\mathbf{n} = \mathbf{n}_1 + \ldots + \mathbf{n}_l$ of $\mathbf{n}$ into sum of positive integers. The dimensions of the representations of this series are equal to(1)

(1) (Added in proof.): There is a very interesting paper of I. A. Green. *Trans. Amer. Math. Soc.* (1955).

$$\frac{(k-1)(k^2-1)\ldots(k^l-1)}{(k^{n_1}-1)\ldots(k^{n_e}-1)}, \tag{7}$$

where k is order of the field under consideration. There are exceptions to this formula, the so-called singular representations. The number of these divided by the total number of representations goes to zero with increasing k.

Now we shall explain how to split into series the principal representations of the arbitrary group G. We combine into one series the principal representations contained in the same $T_\chi(g)$ and with the same multiplicities. It seems that the representations belonging to same series are "contrived in the same manner"; their dimensions coincide; formulae for the characters are written in the same way; when realized they give rise to the same special functions.

Since the principal representations enter in $T_\chi(g)$ of general position with the multiplicity 1, the realization of the representation is comparatively easy. It is sufficient to consider the ring of operators, permutable with the representation $T_\chi(g)$, where χ is a character of general position. This ring is commutative (!), and so it decomposes into a direct sum of complex fields. Each summand defines one of the principal representations of G. It should be noted that the procedure leads to interesting classes of special functions on G. Thus in the case of the unimodular group of second order the one-dimensional components of the ring of operators permutable with $T_\chi(g)$ can be expressed as certain sums known as the Kloosterman sums. It is remarkable that the summation is carried sometimes over a certain "contour" in a quadratic extension of the field. Depending on the set of summation we obtain representations of one or another principal series.

8. Zeta-functions associated with homogeneous space over a finite field

The representations of the group G over a finite field being constructed we have the possibility to define the Zeta-function. Here we restrict ourselves to the Zeta-function associated with principal representations with $\chi \equiv 1$; we shall not carry out the "analytic continuation" to the other series. Let G be a Dickson-Chevalley group over a finite field, let Γ be any of its subgroups and let Z be a maximal nilpotent subgroup. In order to construct the Zeta-function we consider the functions $\varphi(g)$ constant on the cosets of G with respect to Z, i.e. the functions satisfying the equation $\varphi(\zeta g) = \varphi(g)\chi(\zeta)$. The operator M acting in the space of these functions is given by

$$M\varphi = \psi(g) = \sum_{\substack{\gamma \in \Gamma \\ \zeta \in Z}} \varphi(\gamma\zeta g)\chi'(\zeta). \tag{8}$$

Then $\psi(\zeta g) = \psi(g)\chi(\zeta)$. The operator M is permutable with the movement T_{g_0} of the group: $T_{g_0}\varphi(g) = \varphi(gg_0)$. So after the decomposition of the representation into irreducible ones it will be given by a matrix which depends on the "number" of the representation and has order equal to the order of Weyl group. By the definition the matrix is the Zeta-function of group G respect to the subgroup Γ. It depends on the index π of representation which is a multiplicative character of the Cartan subgroup. We shall give the explicit form of this function. For this purpose we construct the function $k(g)$ equal for each g to the number of ways in which g can be written in the form $g = \zeta_1\gamma\zeta_2$, where $\gamma \in \Gamma$ and ζ_1 and ζ_2 are in Z. Each g is of the form

$g = \zeta' \delta\, s\, \zeta''$, where δ is in the Cartan subgroup and s is in the Weyl group; therefore $k(g)$ depends only on δ and s: $k(g) = k(\delta, s)$. The Fourier transform of $k(\delta, s)$ with respect to δ is now the Zeta-function:

$$\zeta_s(\pi) = \sum_{\delta} k(\delta, s)\, \pi(\delta).$$

We shall not write down the form of this function for the representations of other principal series.

9. Reprsentations associated with groups over other fields

We shall not consider the case of groups associated with the field of p-adic numbers which were studied by Mautner and Bruhat. We hope that the indicated constructions are applicable to this case and make it possible to describe all representations.

Very interesting is the problem of studying the representations of the Chevalley groups over the field of algebraic functions over the field of complex numbers. Here it is not evident how to define a representation. Apparently the following definition is the most natural. We require T_g to be an operator defined every where except a manifold of lower dimension. We require T_g to depend continuously on g and to satisfy the equation $T_{g_1} T_{g_2} = T_{g_1 g_2}$ everywhere except a manifold of lower dimension. These groups are examples of infinite-dimensional groups for which the problem of constructing their irreducible representations is quite real and interesting. Also in this case the constructions indicated above give many important types of representations. It is possible, however, to exhibit the examples showing that the construction does not exhaust all degenerate representations.

REFERENCES

[1]. Гельфанд, И. М. & Граев, И. М. И., *Труды Моск. Матем. общ.*, 8 (1959), 323–390.

[2]. Гельфанд, И. М. & Пятецкий-Шациро, И. М. И., *Успехи матем. наук*, XIV 2 (1959), 172–194.

[3]. Гельфанд, И. М. & Фомин, И. С. В., *Успехи матем. наук*, 7:1 (47) (1952), 118–137.

[4]. Selberg, A., *Tata Inst. of Fund. Res., Bombay*, 1960, 147–164.

[5]. Selberg, A., *J. Ind. Math. Soc.*, 20 (1956), 47–87.

[6]. Граев, М. И. *Докл. А. Н. С.С.С.Р.*, 113 (1957), 966–969.

[7]. Selberg, A., *Analytic Functions*, vol. 2, 1960.

[8]. Borel, A. & Harish-Chandra, *Ann. Math.*, 75 3 (1962).

[9]. Гельфанд, И. М. & Наймарк, И. М. А., *Труды матем. ин-та А.Н. С.С.С.Р.*, 36 (1950), 1–288).

[10]. Harish-Chandra, *Proc. Nat. Acad. Sci. USA.*, 45 (1959), 570–573.

[11]. Bargman, *Ann. Math.*, 48 (1947), 568–640.

[12]. Гельфанд, И. М. & Граев, И. М. И., *Изв. А.Н. С.С.С.Р, сер. матем.*, 17 (1953), 189–238.

[13]. Граев, М. И., *ТрудыМоск. матем общ.*, VII (1959).

[14]. Hirzebruch, *Analytic Functions*, vol. 2, 1960.

[15]. Dickson, L., *Linear Groups.*

[16]. Chevalley, C., *Tôhoku Math. J.*, 7 (1955), 14–66.

DIE BEDEUTUNG DES LEVISCHEN PROBLEMS FÜR DIE ANALYTISCHE UND ALGEBRAISCHE GEOMETRIE

Von HANS GRAUERT

Vor nun mehr als 50 Jahren (1910) zeigte E. E. Levi [17], daß jede (reell) 3-dimensionale glatte Hyperfläche H im (komplex) 2-dimensionalen komplexen Zahlenraum $\mathbb{C}^2$, die natürliche Grenze einer holomorphen Funktion ist, gewissen Differentialungleichungen genügen muß. Diese Levischen Differentialbedingungen wurden später von Krzoska [16] auf Dimensionen $n>2$ erweitert und führten schließlich zur Definition des pseudokonvexen Randes und Gebietes. Jedes Holomorphiegebiet, d. h. jedes Gebiet G, zu dem eine in G holomorphe, auf dem Rande ∂G überall singuläre Funktion existiert, ist pseudokonvex. Offen blieb, ob auch jedes pseudokonvexe Gebiet ein Holomorphiegebiet ist. Nachdem in Spezialfällen von Faber [7], Behnke [2], u. a. das Problem im positiven Sinne gelöst worden war, durch Lelong und Oka die plurisubharmonischen Funktionen, ein wichtiges Hilfsmittel, geschaffen worden waren, gelang es Oka 1942 [20], das Problem im Falle $n=2$ zu lösen. 1954 wurde sodann von Oka [21], Bremermann [5], Norguet [19] für $n>2$ die Klasse der pseudokonvexen Gebiete mit der Klasse der Holomorphiegebiete identifiziert. Oka führte dieses sogar für unverzweigte, nicht-schlichte und neuerdings auch für unverzweigte unendliche Gebiete durch. Im Falle verzweigter unendlicher Gebiete wurden Gegenbeispiele gefunden.

§ 1. Komplexe Räume

Inzwischen war der Begriff der *komplexen Mannigfaltigkeit* und der des *komplexen Raumes* aufgestellt worden. Die Definition des komplexen Raumes ist dabei mehrfach verallgemeinert worden. Wir benutzen in diesem Aufsatz die Räume von Serre (man vgl. [10]), die nach der z. Zt. geltenden Terminologie *reduzierte komplexe Räume* heißen, also noch nicht der allgemeinsten Definition entsprechen. Da holomorphe Funktionen und Keime von holomorphen Funktionen auftreten, muß die *Garbentheorie* wesentlich herangezogen werden (als Einführung in die Garbentheorie vgl. [8]).

Definition 1. *Ein Paar $\mathfrak{X}=(X,S)$ heißt ein komplex beringter Raum, wenn folgendes gilt:*

1) *X ist ein topologischer Raum.*

2) *S ist eine Untergarbe von $\mathfrak{C}^-=\mathfrak{C}^-(X)$, der Garbe der Keime von komplexwertigen stetigen Funktionen auf X.*

3) *S ist eine Untergarbe von Ringen und enthält die konstante Untergarbe von $\mathfrak{C}$.*

Anstelle von $\mathfrak{X}$ schreiben wir vielfach auch einfach X. S heißt die Strukturgarbe von X. Durch Beschränkung von S erhält man auch über jeder

offenen Teilmenge $B \subset X$ eine Strukturgarbe. B wird dadurch zu einem komplex beringten Raum. Die Schnittflächen in $\mathfrak{C}$ und mithin erst recht die Schnittflächen in S können als stetige komplexwertige Funktionen gedeutet werden. Sie heißen morphe Funktionen.

Spezielle Beispiele komplex beringter Räume werden durch die analytischen Mengen gegeben (d. s. abgeschlossene Teilmengen in Bereichen B des n-dimensionalen komplexen Zahlenraumes $\mathbf{C}^n$, die sich lokal als simultanes Nullstellengebilde von endlich vielen holomorphen Funktionen darstellen lassen).

Es sei also $A \subset B$ eine analytische Menge. Wir bezeichnen mit $\mathcal{O}$ die Garbe der Keime von lokalen holomorphen Funktionen über B und mit $I \subset \mathcal{O}$ die Garbe der Keime, die auf A verschwinden. Es gilt $I \mid B-A = \mathcal{O} \mid B-A$ und $\mathcal{O}(A) = \mathcal{O}/I \mid A$ kann als Untergarbe von $\mathfrak{C}(A)$ aufgefaßt werden. Das Paar $(A, \mathcal{O}(A))$ ist ein komplex beringter Raum.

Definition 2. *Ein komplex beringter Raum $(X, \mathcal{O})$ heißt ein komplexer Raum, wenn:*

1) *X ein Hausdorffscher Raum ist,*
2) *X lokal isomorph zu einem Paar $(A, \mathcal{O}(A))$ ist.*

Im Falle, daß $(X, \mathcal{O})$ ein komplexer Raum ist, heißen die Schnittflächen in $\mathcal{O}$ natürlich holomorphe Funktionen und $\mathcal{O}$ heißt die Garbe der Keime von holomorphen Funktionen. Jede offene Teilmenge von X ist wieder ein komplexer Raum. Der Begriff der holomorphen Abbildung $\psi: (X, \mathcal{O}_1) \to (Y, \mathcal{O}_2)$ eines komplexen Raumes X in einen anderen Y ist wohldefiniert: Gilt $x \in X$, $y = \alpha(x) \in Y$ und ist $\alpha: X \to Y$ eine stetige Abbildung, so erhält man durch $f \to f \circ \alpha$ einen Homomorphismus α_x^* der Halme $\mathfrak{C}_y(Y) \to \mathfrak{C}_x(X)$. ψ heißt holomorph, wenn für alle Punkte $x \in X$ die Abbildung ψ_x^* stets $\mathcal{O}_{2y}$ in $\mathcal{O}_{1x}$ abbildet, d. h. wenn Keime von holomorphen Funktionen stets in Keime von holomorphen Funktionen übergehen.

Um nun zu den komplexen Mannigfaltigkeiten zurückzukommen, müssen wir die Ringe $\mathcal{O}_x$ näher betrachten. Zunächst ist jeder Halm $\mathcal{O}_x$ eine lokale komplexe Algebra. Ein Punkt $x \in X$ heißt ein regulärer Punkt, wenn die lokale Algebra $\mathcal{O}_x$ regulär ist. Man weiß, daß dieses damit äquivalent ist, daß eine offene Umgebung $U = U(x) \in X$ existiert, die zu einem Bereich $B \subset \mathbf{C}^n$ isomorph ist. Durch jeden Isomorphismus (=biholomorphe Abbildung) $\psi: U \to B$ werden die Koordinaten $z_1, \ldots, z_n$ von B nach U hin übertragen. Zwei auf diese Weise gewonnene Koordinatensysteme hängen durch eine biholomorphe Transformation zusammen und ein Keim einer stetigen Funktion ist genau dann holomorph, wenn er sich in eine Potenzreihe nach den lokalen Koordinaten entwickeln läßt. Ein komplexer Raum, der nur aus regulären Punkten besteht, ist also eine komplexe Mannigfaltigkeit. Man kann leicht zeigen:

Es sei X ein komplexer Raum, $N \subset X$ die Menge der nicht regulären Punkte von X. Dann ist N eine nirgends dichte analytische Teilmenge von X.

Der Bereich $X - N$ ist also eine komplexe Mannigfaltigkeit! Für das Folgende von Bedeutung sind auch die Termini: „normal" und „Dimension". X heißt normal in einem Punkte $x \in X$, wenn der lokale Ring $\mathcal{O}_x$ normal ist, unter der Dimension $d_x(X)$ versteht man die Krullsche Dimension von $\mathcal{O}_x$. Es zeigt sich, daß $d_x = \frac{1}{2} \cdot$(topologische Dimension) ist. Ebenso hat man für „normal" einen Vergleichssatz (=Vergleich mit analytischen Eigenschaften): Ein komplexer Raum ist normal (d.h. normal in allen

Punkten $x \in X$), wenn der Fortsetzungssatz von Riemann gilt: Es sei $U \subset X$ eine offene Teilmenge, $A \subset U$ eine nirgends dichte analytische Menge, f eine in $U-A$ holomorphe beschränkte Funktion. Dann läßt sich f auf ganz U als holomorphe Funktion fortsetzen. Natürlich ist eine komplexe Mannigfaltigkeit in allen ihren Punkten normal.

Für analytische Teilmengen $A \subset X$, die man ja wieder als komplexe Räume auffassen kann, gilt noch:

1) *A ist diskret dann und nur dann, wenn $d_x(A)=0$, $x \in A$.*
2) *A liegt nirgends dicht genau dann, wenn $d_x(A) < d_x(X)$, $x \in A$.*

Nirgends dichte analytische Teilmengen von X werden deshalb auch *niederdimensionale analytische Mengen* genannt.

§ 2. Das Levische Problem für komplexe Räume

Um das Levische Problem auf komplexe Räume zu übertragen, ist es zunächst notwendig, den Begriff des Holomorphiegebietes zu verallgemeinern. Das führt uns zu folgender

Definition 3. *Ein komplexer Raum $(X, \mathcal{O})$ heißt ein Steinscher Raum (oder auch holomorph-vollständiger Raum), wenn folgendes gilt:*

1) *X ist holomorph ausbreitbar:* zu jedem Punkt $x \in X$ gibt es endlich viele in X holomorphe Funktionen $f_1, \ldots, f_k$, derart, daß x isolierter Punkt im Nullstellengebilde $\{f_1 = \ldots = f_k = 0\}$ ist.

2) *X ist holomorph konvex:* Zu jeder unendlichen diskreten Menge $D \subset X$ gibt es eine in X holomorphe Funktion, so daß $f(D)$ unbeschränkt ist.

Aus einem Satz von Thullen folgt unmittelbar: *Ein Gebiet $G \subset \mathbf{C}^n$ ist genau dann Holomorphiegebiet, wenn es ein Steinscher Raum ist.* Wichtig ist auch das Verhalten der *Čechschen Kohomologiegruppen.* Ist X ein Steinscher Raum, so gilt nämlich: $H^\nu(X, \mathcal{O}) = 0$ für $\nu = 1, 2, 3, \ldots$.

Den Begriff der Pseudokonvexität übertragen wir mit Hilfe von plurisubharmonischen Funktionen. Es sei $G \subset \mathbf{C}^n$ ein Gebiet.

Definition 4. *Eine über G zweimal stetig differenzierbare reelle Funktion $p(z)$ heißt (streng) plurisubharmonisch, wenn in jedem Punkte $z \in G$ die Levische Form*

$$L(p) = \sum_{\nu,\mu} \frac{\partial^2 p}{\partial z_\nu \partial \bar{z}_\mu} dz_\nu d\bar{z}_\mu$$

positiv (definit) semidefinit ist.

Die wichtigsten Eigenschaften plurisubharmonischer Funktionen wurden von P. Lelong ermittelt. Der Begriff selbst wurde von ihm auf halbstetige Funktionen ausgedehnt. — Es sei nun X ein komplexer Raum:

Definition 5. *Eine reelle Funktion $p(x)$ über X heißt (streng) plurisubharmonisch, wenn es zu jedem Punkt $x \in X$ eine Umgebung $U = U(x)$, ein Gebiet $G \subset \mathbf{C}^n$, eine analytische Menge $A \subset G$, eine biholomorphe Abbildung ψ: $U \xrightarrow{\sim} A$ und eine in G (streng) plurisubharmonische Funktion $\hat{p}$ gibt, so daß $p \mid U = \hat{p} \circ \psi$ ist.*

In analoger Weise kann man den Begriff der k-mal stetig differenzierbaren Funktion auf X übertragen. — Pseudokonvexe Gebiete werden nun auf folgende Weise definiert:

DEFINITION 6. *Ein relativ kompaktes Teilgebiet G eines komplexen Raumes X heißt (streng) pseudokonvex, wenn es zu jedem Randpunkt $x_0 \in \partial G$ eine Umgebung $U = U(x_0)$ und eine in U (streng) plurisubharmonische Funktion p gibt, so daß*

$$U \cap G = \{x \in U : p(x) < p(x_0)\} \text{ gilt.}$$

Man zeigt leicht:

Ist G streng pseudokonvex, so kann man sogar eine Umgebung $U = U(\partial G)$ und eine in U streng plurisubharmonische Funktion p finden, derart, daß $U \cap G = \{x \in U : p(x) < 0\}$ (vgl. [13]).

Durch Anwendung eines funktionalanalytischen Satzes von L. Schwartz [24] zeigt man nun (vgl. [11]):

Ist $G \Subset X$ streng pseudokonvex, so gilt $\dim H^\nu(G, \mathcal{O}) < \infty$, $\nu = 1, 2, 3, \ldots$. *Ab einem ν_0 verschwindet außerdem $H^\nu(G, \mathcal{O})$.*

Aus diesem grundlegenden Resultat, das ohne größere Schwierigkeit hergeleitet werden kann, folgt sofort, daß G, falls streng pseudokonvex, auch holomorph konvex ist und weiter, daß G ein Steinscher Raum ist, sofern es keine kompakten analytischen Teilmengen A mit $\dim A > 0$ enthält (in einem Steinschen Raum gibt es natürlich niemals Teilmengen dieser Art).

Ein Gebiet $G \subset \mathbf{C}^n$ ist stets holomorph-ausbreitbar, wie man unter Benutzung der Koordinatenfunktionen $z_1, \ldots, z_n$ sofort einsieht, und deshalb holomorph konvex genau dann, wenn es ein Holomorphiegebiet ist. Wir haben also das Levische Problem für (streng pseudokonvexe) Gebiete $G \subset \mathbf{C}^n$ aufs Neue gelöst und gleichzeitig auf komplexe Räume verallgemeinert. Der Beweis ist übrigens wesentlich einfacher als der ursprüngliche, der von Oka angegeben wurde und hauptsächlich auf einer geschickten Anwendung von Integralen beruht. Es bleibt zu erwähnen, daß die bei uns angewendeten funktionalanalytischen Methoden ebenso „konstruktiv" sind wie die Integration.

Es gibt pseudokonvexe Gebiete $G \Subset X$, die nicht holomorph-konvex sind und für die etwa $\dim H^1(G, \mathcal{O}) = \infty$ gilt.

Der Begriff der Pseudokonvexität ist durch R. Narasimhan von relativ kompakten Teilgebieten auf beliebige komplexe Räume verallgemeinert worden. Ein komplexer Raum X heißt streng pseudokonvex, wenn es eine kompakte Teilmenge $K \subset X$ und eine in $X - K$ streng plurisubharmonische Funktion p gibt, so daß für jedes $r > 0$ die Menge $\{x \in X - K : p(x) < r\} \Subset X$. Ist $G \Subset X$ und streng pseudokonvex im Sinne von Definition 6, so ist G auch ein streng pseudokonvexer komplexer Raum. R. Narasimhan hat nun unter Verwendung eines neuen Approximationssatzes gezeigt [18]:

SATZ 1. *Es sei X streng pseudokonvex. Dann ist X holomorph-konvex und es gilt* $\dim H^\nu(X, \mathcal{O}) < \infty$, $\nu \geqslant 1$, *und* $H^\nu(X, \mathcal{O}) = 0$, $\nu \geqslant \nu_0$.

Eine Charakterisierung der Steinschen Räume gelingt durch die Pseudovollständigkeit:

DEFINITION 7. *Ein komplexer Raum X heißt pseudovollständig, wenn es*

eine auf X streng plurisubharmonische Funktion p gibt, so daß für jedes $r>0$ die Menge $\{x\in X: p(x)<r\}$ relativ-kompakt in X liegt.

R. Narasimhan hat gezeigt:

SATZ 2. *X ist ein Steinscher Raum genau dann, wenn X pseudovollständig ist.*

Von W. Rothstein wurde der Begriff der q-Konvexität in die Literatur eingeführt und bei den Kontinuitätssätzen für analytische Mengen mit viel Erfolg verwendet. Wir definieren hier zunächst q-konvexe Funktionen. Es sei wieder $G\subset \mathbf{C}^n$ ein Gebiet und $p(z)$ eine reelle zweimal stetig differenzierbare Funktion über G.

DEFINITION 8. *p heißt q-konvex über G, wenn in jedem Punkt $z\in G$ die Levische Form $L(p)$ mindestens $n-q+1$ positive Eigenwerte hat (mit $q=1,2,3,\ldots,n$).*

Die 1-konvexen Funktionen fallen also mit den streng plurisubharmonischen Funktionen zusammen. Jede q-konvexe Funktion ist erst recht $(q+1)$-konvex.

Es sei nun X ein komplexer Raum und p eine über X reelle Funktion.

DEFINITION 9. *p heißt q-konvex über X, wenn es zu jedem Punkt $x\in X$ eine Umgebung $U=U(x)$, ein Gebiet $G\subset \mathbf{C}^n$, eine analytische Menge $A\subset G$, eine biholomorphe Abbildung $\psi: U\xrightarrow{\sim} A$ und eine in G q-konvexe Funktion $\hat{p}$ gibt, so daß $\hat{p}\circ\psi=p\,|\,U$ gilt.*

Entsprechend der Pseudokonvexität nennen wir einen komplexen Raum X *q-konvex* (*q-vollständig*), wenn es eine kompakte Teilmenge $K\subset X$ und in $X-K$ (und in X) eine q-konvexe Funktion p gibt, so daß jede Menge $\{p(x)<r\}$ relativ kompakt in X liegt.

In einer gemeinsamen Arbeit mit A. Andreotti wurde gezeigt [1]:

SATZ 3. *Ist X q-konvex, so gilt* $\dim H^\nu(X,\mathcal{O})<\infty$ *für $\nu\geqslant q$ und $H^\nu(X,\mathcal{O})=0$, $\nu\geqslant\nu_0$. Setzt man noch voraus, daß X q-vollständig ist, so folgt sogar $H^\nu(X,\mathcal{O})=0$, $\nu\geqslant q$.*

Der Beweis wurde in [1] erbracht. Es werden zunächst relativ-kompakte Teilgebiete $G\Subset X$ untersucht. Ein solches Gebiet heißt q-konvex, wenn es eine Umgebung $U=U(\partial G)\subset X$ und eine in U q-konvexe Funktion p gibt, so daß $G\cap U=\{x\in U: p(x)<0\}$. Man zeigt sodann, daß es zu jedem Randpunkt $x\in\partial G$ beliebig kleine Umgebungen $V=V(x)$ gibt, derart, daß $H^\nu(V,\mathcal{O})=0$ für $\nu\geqslant q$. Deshalb folgt: Ist $\varepsilon>0$ hinreichend klein gewählt, so gilt für $G'=G\cup\{x\in U: p(x)<\varepsilon\}$, daß der natürliche Homomorphismus: $H^\nu(G',\mathcal{O})\to H^\nu(G,\mathcal{O})$ bijektiv ist für $\nu\geqslant q$. Aus „surjektiv" und dem Lemma von Schwartz ergibt sich sodann: $\dim H^\nu(G,\mathcal{O})<\infty$ für $\nu\geqslant q$. Weiter wird ein Approximationssatz bewiesen. In der Menge der Kozyklen $Z'=Z^\nu(U',\mathcal{O})$, $Z=Z^\nu(U,\mathcal{O})$ (bzgl. geeigneter Überdeckungen von G' bzw. G) läßt sich eine natürliche Topologie einführen, die Z',Z zu Fréchetschen Räumen macht. Die kanonische Abbildung $Z'\to Z$ liefert eine in Z dichte Bildmenge, wenn $\nu\geqslant q-1$ ist. Jeder Kozyklus über G läßt sich also beliebig stark durch Kozyklen über G' approximieren. Mit Hilfe dieser Approximationsaussage folgt leicht, daß $H^\nu(X,\mathcal{O})\to H^\nu(G,\mathcal{O})$, $\nu\geqslant q$ bijektiv ist, wenn $G=\{x\in X-K: p(x)<r\}\cup K$ gesetzt wird. $K\subset X$ ist dabei eine kompakte Menge, $p(x)$ eine in $X-K$ q-konvexe Funktion, für r und p gelte: $r>p(x)$, wenn $x\in U(\partial K)\cap$

$(X-K)$ (mit $U(\partial K)$ = eine kleine Umgebung). Man gewinnt also aus den Endlichkeitsaussagen für G die entsprechenden Aussagen für X. Im Falle, daß X q-vollständig ist, kann man K = (leere Menge) setzen und erhält: $H^{\nu}(X, \mathcal{O}) = 0$, $\nu \geqslant q$. Damit ist der Beweis vollständig.

Der Satz bleibt richtig, wenn man $\mathcal{O}$ durch eine beliebige kohärente analytische Garbe über X ersetzt. Auf den Begriff der kohärenten Garbe soll hier jedoch nicht näher eingegangen werden.

Die Lösung des Levischen Problems für komplexe Mannigfaltigkeiten ist inzwischen auch auf andere Weise versucht worden. Mr. Nishino (Nara) hat sich sehr eng an den Gedankengang von Oka angeschlossen und mit Hilfe der Okaschen Integralmethode u. a. gezeigt:

Eine komplexe Mannigfaltigkeit X ist genau dann eine Steinsche Mannigfaltigkeit, wenn es eine in X streng plurisubharmonische Funktion $p(x)$ gibt, für die jede Menge $\{p(x) < r\}$ relativ kompakt in X liegt.

Der Begriff „streng plurisubharmonisch" wird dabei allgemeiner gebraucht, die Funktion $p(x)$ braucht nicht notwendig zweimal stetig differenzierbar zu sein. U. a. fallen bei unverzweigten Gebieten G über dem P^n gewisse Funktionen $p(x) = -lg\ \delta(x)$ darunter, wobei $\delta(x)$ einen (nichteuklidischen) Abstand der Punkte $x \in G$ vom Rande ∂G bezeichnet. Ist ∂G pseudokonvex, so gelingt es jedenfalls, eine in G streng plurisubharmonische Funktion p zu konstruieren mit $\{p(x) < r\} \leqslant G$. In Nara kann man deshalb zeigen:

Ein unverzweigtes, nicht kompaktes Gebiet über dem P^n ist eine Steinsche Mannigfaltigkeit genau dann, wenn es einen pseudokonvexen Rand hat.

Es ist dem Verfasser unbekannt, ob und wo dieses Resultat veröffentlicht worden ist.

§ 3. Exzeptionelle analytische Mengen

Wir werden uns im folgenden mit der Anwendung der Levischen Aussage in der analytischen und algebraischen Geometrie befassen. Zunächst seien exzeptionelle analytische Mengen betrachtet. Zu dem Zwecke werde zunächst der Begriff der *Modifikation* von komplexen Räumen untersucht.

Modifikationen traten zum ersten Male 1950 in einer gemeinsamen Arbeit von H. Behnke und K. Stein auf [3]. Die beiden Autoren verstanden darunter einen Prozeß, der es gestattet, einen komplexen Raum X in einer Teilmenge N abzuändern. Die Teilmenge N wird herausgenommen und durch eine Punktmenge N' ersetzt, so daß sich die komplexe Struktur von $X-N$ nach $X' = (X-N) \cup N'$ fortsetzen läßt. Man hat die identische Abbildung $\psi: X-N \to X'-N'$. Da man komplexe Räume, die man biholomorph aufeinander abgebildet hat, als gleich ansehen kann, definiert man:

Definition 10. *Eine Modifikation ist ein Quintupel (X', N', ψ, N, X), wobei:*

1) *X', X komplexe Räume,*

2) *$N \subset X$, $N' \subset X'$ abgeschlossene Teilmengen,*

3) *$\psi: X'-N' \xrightarrow{\sim} X-N$ eine biholomorphe Abbildung sind, derart, daß folgendes gilt:*

1) Ist $U = U(N) \subset X$ eine Umgebung von N, so ist $U' = \psi^{-1}(U-N) \cup N'$ eine Umgebung von N'.

2) N ist dünn, d. h. zu jedem Punkt $x_0 \in N$ gibt es eine Umgebung $U = U(x_0)$ und eine in U holomorphe, nirgends identisch verschwindende Funktion f, so daß $U \cap N \subset \{x \in U : f(x) = 0\}$.

Modifikationen können sehr pathologisch sein, wie bereits H. Behnke und K. Stein gezeigt haben. Man ging deshalb schon bald dazu über, spezielle Klassen zu untersuchen. Eine solche Klasse bilden die eigentlichen Modifikationen, die in [9] eingeführt wurden:

DEFINITION 11. *Eine Modifikation (X', N', ψ, N, X) heißt eigentlich, wenn sich ψ zu einer eigentlichen holomorphen Abbildung $\hat{\psi} : X' \to X$ fortsetzen läßt.*

$\hat{\psi}$ ist natürlich eindeutig bestimmt. Es folgt sofort, daß $\hat{\psi}$ surjektiv ist. Wählt man N' und N so klein wie möglich, so sind N', N analytische Mengen und unter der zusätzlichen Voraussetzung, daß X', X normale komplexe Räume sind, hat die Menge N sogar in jedem ihrer Punkte die Codimension 2, während aus der Annahme, daß X' normal und X eine komplexe Mannigfaltigkeit ist, folgt, daß N' überall 1-codimensional ist.

Spezielle eigentliche Modifikationen wurden schon sehr früh untersucht. Sind X', X vollständige normale algebraische Räume (Varietäten), so ist (X', N', ψ, N, X) genau dann eine eigentliche Modifikation, wenn $\hat{\psi} : X' \to X$ eine reguläre, birationale Transformation ist. Aus der algebraischen Geometrie wurde auch der σ-Prozeß (monoidale Transformation) in die komplexe Analysis übernommen. Man versteht darunter einen Prozeß, der es gestattet, einen regulären Punkt durch eine höher dimensionale analytische Menge zu ersetzen. Er ist eine eigentliche Modifikation (X', N', ψ, N, X), bei der N ein regulärer Punkt, $N' = P^{n-1}$ (mit $n = \dim_{\mathbb{C}} X$) eine singularitätenfrei eingebettete analytische Menge und X' in der Umgebung von N' eine komplexe Mannigfaltigkeit ist. Komplexe Mannigfaltigkeiten werden also durch den σ-Prozeß stets in komplexe Mannigfaltigkeiten überführt. Es gibt andere eigentliche Modifikationen, bei denen ein Punkt x_0 durch eine höherdimensionale analytische Menge A ersetzt wird. Ist x_0 regulär, so wird jedoch jetzt nicht jeder Punkt von A regulär sein. Wir beschäftigen uns in diesem Paragraphen mit Modifikationen dieser Art in umgekehrter Richtung und nennen eine analytische Menge exzeptionell, wenn sie durch eine eigentliche Modifikation durch einen Punkt ersetzt werden kann. Um den Begriff exakt einführen zu können, definieren wir zunächst:

DEFINITION 12. *Eine analytische Teilmenge A heißt nirgends diskret, wenn sie keine isolierten Punkte enthält.*

Dimensionstheoretisch gesehen ist A genau dann nirgends diskret, wenn für alle Punkte $x \in A$ gilt: $d_x(A) > 0$. Es sei fortan $A \subset X$ eine nirgends diskrete niederdimensionale, kompakte analytische Teilmenge.

DEFINITION 13. *A heißt exzeptionell, wenn es einen komplexen Raum Y, eine diskrete Menge $D \subset Y$ und eine holomorphe eigentliche Abbildung $\psi : X \to Y$ gibt, so daß folgendes gilt:*

1) *$\psi(A) = D$,*

2) *ψ bildet $X - A$ biholomorph auf $Y - D$ ab,*

3) *Ist $U = U(D)$ eine offene Umgebung, f eine holomorphe Funktion in $V = \psi^{-1}(U)$, so gibt es eine in U holomorphe Funktion $\mathfrak{f}$ mit $f = \mathfrak{f} \circ \psi$.*

Die Eigenschaft 3 bedeutet dabei, daß in der Nähe von D hinreichend viele holomorphe Funktionen existieren, die komplexe Struktur von Y dort also nicht zu dünn ist. Ist A exzeptionell, so ist das Paar (ψ, Y) bis auf eine naheliegende Äquivalenz eindeutig bestimmt, d. h. sind (ψ_ν, Y_ν), $\nu = 1,2$ zwei Paare mit den Eigenschaften 1–3, so gibt es eine biholomorphe Abbildung $\alpha: Y_1 \to Y_2$, so daß das Diagramm:

$$\begin{tikzcd} & X \arrow[dl, "\psi_1"'] \arrow[dr, "\psi_2"] & \\ Y_1 \arrow[rr, "\alpha"] & & Y_2 \end{tikzcd}$$

kommutativ ist. Der Beweis dieser Aussage ist nicht schwer. Übrigens: Ist X ein normaler komplexer Raum, so ist Y wieder normal, wie man durch die analytische Charakterisierung von „normal" sofort einsieht. Dagegen braucht Y keine komplexe Mannigfaltigkeit zu sein, auch dann nicht, wenn X es ist. Es sei $A = A_1 \cup \ldots \cup A_k$ die Zerlegung von A in zusammenhängende Komponenten. Wir setzen $d_\nu = \psi(A_\nu)$. Es gilt dann $d_\nu \neq d_\mu$ für $\nu \neq \mu$ und $D = \{d_1, \ldots, d_k\}$.

Wir sagen auch, daß eine exzeptionelle analytische Menge A „niedergeblasen" werden kann. Im allgemeinen ist eine analytische Menge nicht exzeptionell. Ist X homogen, so gibt es keine kompakte, exzeptionelle analytische Teilmenge in X. Man kann sogar zeigen: In einer hinreichend kleinen Umgebung einer exzeptionellen Menge gibt es niemals eine weitere.

Die Frage, die sich stellt, lautet also: *Wann ist eine kompakte, niederdimensionale, nirgends diskrete analytische Menge $A \subset X$ exzeptionell?* Es zeigt sich nun, daß man mit Hilfe der Levischen Aussage leicht zu einem einfachen Kriterium kommen kann. Dazu ist es notwendig, die Remmertsche Reduktionstheorie anzuwenden. Diese Theorie, die zu vielen wichtigen Ergebnissen führt, wurde 1957 [22] von R. Remmert aufgestellt. Ihre Resultate wurden später [6] von H. Cartan verschärft. Die Beweise benutzen wesentlich die Sätze von K. Stein über analytische Zerlegungen. Vgl. auch die Methoden von H. Rossi [23].

SATZ 4. *Es sei X ein holomorph-konvexer komplexer Raum. Dann gibt es einen Steinschen Raum Y und eine eigentliche holomorphe Abbildung $\psi: X \to Y$, so daß gilt: Ist $U \subset Y$ eine offene Menge, f eine holomorphe Funktion über $V = \psi^{-1}(U)$, so gibt es eine über U holomorphe Funktion $\mathfrak{f}$ mit $f = \mathfrak{f} \circ \psi$.*

Das Paar (ψ, Y) ist wieder bis auf naheliegende Äquivalenz eindeutig bestimmt. Die Fasern $X_y = \psi^{-1}(y)$, $y \in Y$, sind kompakte analytische Teilmengen von X und — das ist im allgemeinen Fall sehr tiefliegend — sie sind zusammenhängend.

Ist $y_0 \in Y$ und $x_0 = \psi^{-1}(y_0)$ ein einzelner Punkt, so gibt es Umgebungen von y_0 und x_0, die durch ψ biholomorph aufeinander abgebildet werden. Falls X ein normaler komplexer Raum ist, folgt wieder wie vorhin, daß auch Y normal ist.

Es dürfte jetzt dem Leser einleuchten, wie man das Resultat von Cartan und Remmert anzuwenden hat. Ist X ein komplexer Raum, so heißt eine kompakte, nirgends diskrete analytische Teilmenge $A \subset X$ maximale kompakte analytische Teilmenge von X, wenn jede beliebige andere nirgends diskrete, kompakte analytische Teilmenge von X in A enthalten ist. Im allgemeinen besitzt ein komplexer Raum X natürlich keine maximale

kompakte analytische Teilmenge — man nehme etwa als Beispiel das karte sische Produkt des Einheitskreises der z-Ebene mit dem komplex projekti ven Raum P^n. Jedoch: existiert A, so ist A auch eindeutig bestimmt.

Unter Verwendung der Remmertschen Reduktionstheorie und der Le vischen Aussage zeigt man nun:

SATZ 5. *Es sei X ein streng pseudokonvexer komplexer Raum. Dann besitzt eine maximale kompakte analytische Teilmenge A* [13].

Um das gesuchte Kriterium für exzeptionelle analytische Mengen z gewinnen, bezeichne jetzt X einen beliebigen komplexen Raum, $A \subset X$ se eine niederdimensionale, nirgends diskrete, kompakte analytische Tei menge. Es sei $T = T(A)$ eine relativ-kompakte streng pseudo-konvex Umgebung von A und A sei maximale kompakte analytische Teilmeng von T. Da nach der Levischen Aussage T holomorph-konvex ist, könne wir die Remmertsche Reduktionstheorie heranziehen. Wir erhalten eine komplexen Raum T^* und eine eigentliche holomorphe Abbildung $\tilde{\psi}: T \to T^*$ die alle Eigenschaften aus Satz 4 besitzt. Der berühmte Projektionssat von Remmert besagt nun, daß jedes Bild einer analytischen Menge unte einer eigentlichen und holomorphen Abbildung wieder eine analytisch Menge ist. $\tilde{\psi}(A)$ ist also analytische Teilmenge von T^* und da $\tilde{\psi}$ stetig is ist sie auch wie A kompakt. In einem Steinschen Raum liegt aber jed kompakte analytische Menge diskret. Wir schreiben $\tilde{\psi}(A) = D$. Es folg daß $\tilde{\psi}^{-1}(y)$ für $y \in T^* - D$ einpunktig ist. Anderenfalls wäre $\dim \tilde{\psi}^{-1}(y) >$ und damit $\tilde{\psi}^{-1}(y) \in A$ und $y \in D$. Also wird $T - A$ durch $\tilde{\psi}$ biholomorph au $T^* - D$ abgebildet. Durch Zusammenheften von T bzw. T^* mit $X -$ erhält man schließlich den gewünschten Raum $Y = T^* \cup (X - T)$ und di eigentliche holomorphe Abbildung $\psi: X \to Y$ mit den verlangten Eiger schaften. A ist also exzeptionell. — Umgekehrt kann man bei exzeptionelle Mengen leicht eine Umgebung T konstruieren.

SATZ 6. *Eine niederdimensionale, nirgends diskrete kompakte analytisch Teilmenge A eines komplexen Raumes X ist exzeptionell genau dann, wenn eine offene streng pseudokonvexe Umgebung $T = T(A) \leqslant X$ gibt, derart, da A die maximale kompakte analytische Teilmenge von T ist* [13].

In der algebraischen Geometrie ist die entsprechende Aussage falsch. Is X ein kompakter (=vollständiger) algebraischer Raum, $A \subset X$ eine exzep tionelle analytische Menge, so braucht der Raum Y, den man durch Niede blasen von A erhält, keineswegs mehr algebraisch zu sein. Das einfachst Beispiel dafür wurde von Hironaka angegeben. Man nehme etwa den zwe dimensionalen komplex projektiven Raum T. $D \subset T$ sei eine singularitäter freie abgeschlossene algebraische Fläche 3. Ordnung. Wir blasen durch Prozesse 10 geeignete Punkte $x_r \in D$ auf. Das Resultat ist eine 2-dimer sionale kompakte komplexe Mannigfaltigkeit T^*, die — wie man zeigt - projektiv algebraisch ist. D wird durch die σ-Prozesse in eine singularitäter freie, irreduzible Teilmenge D^* transformiert. D^* ist jetzt exzeptionel Durch Niederblasen erhält man einen 2-dimensionalen kompakten kor plexen Raum Y mit folgenden Eigenschaften:

1) Y ist normal.
2) Y hat nur einen singulären Punkt.

3) Y besitzt 2 analytisch und algebraisch unabhängige meromorphe Funktionen.

4) Y ist trotzdem nicht algebraisch (nicht einmal im allgemeinen Weilschen Sinne).

§4. Geradenbündel und Hodge'sche Räume

Wir werden in diesem Abschnitt eine Anwendung der Levischen Aussage auf komplex-analytische Geradenbündel über kompakten komplexen Räumen geben (man vgl. [13]). Es sei also X ein kompakter komplexer Raum, F sei ein weiterer komplexer Raum, der durch eine holomorphe Abbildung π auf X abgebildet ist. Jede Faser $F_x = \pi^{-1}(x)$, $x \in X$ trage außerdem noch die Struktur eines eindimensionalen komplexen Vektorraumes. Es gilt also $F_x \simeq \mathbf{C}$. Gibt es nun außerdem noch zu jedem Punkt $x_0 \in X$ eine Umgebung $U = U(x_0)$ und eine biholomorphe Abbildung $V = \pi^{-1}(U) \xrightarrow{\sim} U \times \mathbf{C}$, die jede Faser F_x, $x \in U$ auch als komplexe Vektorräume isomorph auf $x \times \mathbf{C}$ abbildet, so heißt $F \xrightarrow{\pi} X$ ein komplexanalytisches Geradenbündel über X. Wie üblich verstehen wir unter einer Schnittfläche in F eine holomorphe Abbildung $\psi: X \to F$ mit $\pi \circ \psi =$ Identität $x \to x$. Ordnet man jedem Punkt $x \in X$ das Nullelement $O_x \in F_x$ zu, so erhält man die Nullschnittfläche $\mathfrak{O} = \mathfrak{O}(F)$. Wir bezeichnen die Bildmenge von $\mathfrak{O}$ im folgenden auch mit $\mathfrak{O}$, also $\mathfrak{O} = \{O_x\}$.

Ersetzen wir jeden komplexen Vektorraum F_x durch seinen dualen (bzw. durch das k-fache Tensorprodukt mit sich selbst), so erhalten wir das duale komplex-analytische Geradenbündel F^* (bzw. das komplex-analytische Geradenbündel F^k).

Es seien $s_0, \ldots, s_k$ Schnittflächen in F über X. In jedem Punkt $x \in X$ gelte für wenigstens ein ν die Ungleichung $s_\nu(x) \neq O \in F_x$. Ist dann $\alpha: F_x \xrightarrow{\sim} \mathbf{C}$ ein Isomorphismus, so ist $(\alpha s_0(x), \ldots, \alpha s_k(x)) \neq 0$ ein $(k+1)$-tupel komplexer Zahlen. Da α bis auf einen multiplikativen Faktor bestimmt ist, erhalten wir zu x genau einen Punkt $\psi(x) \in P^k$. ψ ist eine holomorphe Abbildung $X \to P^k$.

Definition 14. *F heißt reich (ample), wenn es eine natürliche Zahl n und holomorphe Schnittflächen $s_0, \ldots, s_k$ in F^n gibt, so daß ψ eine biholomorphe Einbettung von X in den P^k ist.*

Existiert ein reiches Geradenbündel über X, so ist X natürlich projektiv algebraisch. Ferner stehen die Äquivalenzklassen von Einbettungen von X in die komplex projektiven Räume mit den Äquivalenzklassen von reichen Geradenbündeln auf X in Beziehung. Es ist deshalb von besonderer Wichtigkeit zu wissen, wann ein Geradenbündel reich ist.

Im Falle, daß X eine kompakte komplexe Mannigfaltigkeit ist, wurde ein einfaches Kriterium von Kodaira angegeben [15]. Es sei $c(F)$ das Bild der 1. Chernschen Klasse von F in der zweiten Kohomologiegruppe $H^2(X, R)$ von X mit Koeffizienten in den reellen Zahlen R. Nach dem Satz von De Rham ist jede Kohomologieklasse aus $H^2(X, R)$ cohomolog zu einer geschlossenen, reellen Differentialform der Dimension 2.

Satz 7. (von Kodaira). *F ist reich dann und nur dann, wenn $c(F)$ cohomolog zu einer geschlossenen, reellen Differentialform $\varphi = i \cdot \sum_{\nu,\mu} g_{\nu\mu} dz_\nu \wedge d\bar{z}_\mu$*

ist, bei der die zugeordnete Hermitesche Form $\sum g_{\nu\mu} dz_\nu \cdot d\bar{z}_\mu$ *in jedem Punkte positiv definit ist.*

Dabei bezeichnen natürlich $z_1, \ldots, z_n$ lokale Koordinaten in X. Der Beweis des Satzes benutzt wesentlich Potentialtheorie und kann deshalb nicht ohne weiteres auf komplexe Räume übertragen werden. Es war deshalb wünschenswert, für komplexe Räume ein anderes Kriterium zu gewinnen. Es sei also nun wieder X ein beliebiger kompakter komplexer Raum, F ein komplex-analytisches Geradenbündel über X.

SATZ 8. *F ist genau dann reich, wenn die Nullschnittfläche $\mathfrak{O} \subset F^*$ exzeptionell ist.*

Der Beweis dieses Satzes ist algebraischer Natur und wurde von Grothendieck auch in der algebraischen Geometrie durchgeführt [14].

Nun ist bei einem Geradenbündel in jeder relativ kompakten, streng pseudokonvexen Umgebung T um die Nullschnittfläche $\mathfrak{O}$ die Menge $\mathfrak{O}$ selbst die maximale kompakte analytische Teilmenge. Es folgt deshalb in Verbindung mit dem Kriterium für exzeptionelle analytische Mengen:

SATZ 9. *F ist genau dann reich, wenn es eine relativ kompakte, streng pseudokonvexe Umgebung $T = T(\mathfrak{O}) \leqslant F^*$ gibt.*

Ein Geradenbündel F^*, bei dem $T(\mathfrak{O})$ existiert, heißt auch ein negatives Geradenbündel. Existiert bei F^* die Umgebung $T(\mathfrak{O})$, so nennen wir F selbst positiv.

Der Begriff der Kählerschen Metrik läßt sich in jedem komplexen Raum X definieren. Es sei $N \subset X$ die analytische Menge der singulären Punkte von X. Da N nirgends dicht in X liegt, ist folgende Definition sinnvoll.

DEFINITION 15. *Eine Hermitesche Metrik* $ds^2 = \sum g_{\nu\mu} d\bar{z}_\nu dz_\mu$ *(mit stetigen Koeffizienten) in $X - N$ heißt eine Kählersche Metrik in X, wenn für die zugeordnete äußere Form* $\varphi = i \cdot \sum_{\nu,\mu} g_{\nu\mu} dz_\nu \wedge d\bar{z}_\mu$ *folgendes gilt: Zu jedem Punkt $x \in X$ gibt es eine Umgebung $U = U(x)$ und eine in U streng plurisubharmonische Funktion p, so daß $d'd''p \mid U - N = \varphi \mid U - N$ ist.*

Dabei bezeichnet d' die totale Ableitung nach den lokalen Koordinaten z_ν und d'' die totale Ableitung nach den $\bar{z}_\nu$. Bei einer Kählerschen Metrik ist also gefordert, daß lokal stets ein Potential p existiert.

Wie bei komplexen Mannigfaltigkeiten definiert nach dem De-Rhamschen Satz die Form φ eine Kohomologieklasse $\xi \in H^2(X, R)$. Ist X kompakt und liegt ξ im Bild der zweiten Kohomologiegruppe von X mit ganzzahligen Koeffizienten, so nennen wir ds^2 eine Hodgesche Metrik. Ein komplexer Raum, auf dem es eine Hodgesche Metrik gibt, heißt ein Hodgescher Raum.

Bei einem normalen komplexen Raum kann man nun zu einer Hodgeschen Metrik leicht ein komplex-analytisches Geradenbündel F mit $c(F) = \xi$ konstruieren. Für $\mathfrak{O} \subset F^*$ weist man nach: Es gibt eine streng pseudokonvexe Umgebung $T = T(\mathfrak{O}) \leqslant F^*$. F ist also reich und X projektiv algebraisch.

SATZ 10. *Jeder Hodgesche Raum ist projektiv algebraisch.*

Damit ist der berühmte Satz von Kodaira auf normale komplexe Räume übertragen worden. Die Methoden sind dabei von den Kodairaschen völlig verschieden und beruhen einzig und allein auf der Levischen Aussage

Diese führt also auch bei kompakten komplexen Räumen zu wichtigen Resultaten.

Für zweidimensionale kompakte komplexe Mannigfaltigkeiten X gewinnt man ein sehr einfaches Kriterium dafür, daß X projektiv algebraisch ist. Es gilt nämlich folgender in dieser Allgemeinheit zuerst von van de Ven bewiesene:

SATZ 11. *X ist genau dann projektiv algebraisch, wenn es eine analytische Menge $A \subset X$ gibt, die in die irreduziblen Komponenten $A_1, \ldots, A_k$ zerfällt, so daß die Schnittmatrix $(a_{\nu\mu})$ mit $a_{\nu\mu} = (A_\nu, A_\mu)$ weder negativ definit noch negativ semidefinit ist.*

Um zu zeigen, daß das Kriterium hinreichend ist, zeigt man zunächst daß die Voraussetzung: „Mindestens ein Eigenwert von $(a_{\nu\mu})$ positiv" damit äquivalent ist, daß es natürliche Zahlen $c_1, \ldots, c_k > 0$ gibt, so daß auch $d_\nu = \sum a_{\nu\mu} c_\mu > 0, \nu = 1, \ldots, k$, gilt (genauer: Nachdem man evtl. zu einer quadratischen Untermatrix von $(a_{\nu\mu})$ übergegangen ist.) Die Schnittzahl von A_ν mit dem Zyklus $\mathfrak{A} = \sum c_\mu A_\mu$ ist gleich d_ν und deshalb, wenn das Kriterium erfüllt ist, stets positiv. Daraus folgt dann: Die Beschränkung des Geradenbündels F, das zu dem Divisor $\mathfrak{A}$ gehört, auf A ist positiv. Unter Verwendung dieses Zwischenresultates zeigt man weiter: $X - A$ ist streng pseudokonvex. Es sei B die maximale kompakte analytische Teilmenge von $X - A$. B zerfalle in die irreduziblen Komponenten $B_1, \ldots, B_m$. Man kann nun natürliche Zahlen $e_1, \ldots, e_m$ finden, so daß das zu dem Divisor $\sum c_\mu A_\mu - \sum e_\nu B_\nu$ gehörende Geradenbündel über ganz X reich ist, wobei man unser Kriterium für reiche Geradenbündel benutzen muß. Damit hat man bewiesen, daß X projektiv algebraisch ist. Wieder einmal wurde also die Levische Aussage wesentlich herangezogen.

Eine analytische Menge $A \subset X$ mit einer nicht negativ semidefiniten Schnittmatrix gibt es stets, wenn auf X zwei unabhängige meromorphe Funktionen existieren. Also ist in unserem Satz das bekannte Resultat von Chow und Kodaira enthalten, daß jede kompakte, 2-dimensionale komplexe Mannigfaltigkeit mit zwei unabhängigen meromorphen Funktionen projektiv algebraisch ist.

§ 5. Normalenbündel

Um das Kriterium für exzeptionelle analytische Mengen aus § 3 anwenden zu können, ist es notwendig, Kenntnis der Struktur einer ganzen Umgebung von A zu haben. Wir wollen zeigen, daß das Normalenbündel von A bereits entscheidend ist, und dieses kann natürlich bedeutend leichter bestimmt werden.

Zunächst betrachten wir komplex-analytische Vektorraumbündel über kompakten komplexen Räumen X. Darunter versteht man einen komplexen Raum V, der durch eine holomorphe Abbildung $\pi: V \to X$ auf X bezogen ist. Die Fasern $V_x = \pi^{-1}(x)$ tragen außerdem die Struktur eines q-dimensionalen komplexen Vektorraumes und lokal ist wieder alles zu $U \times \mathbf{C}^q$, $U = U(x) \subset X$, isomorph. Die Zahl q ist dabei konstant und hängt nicht von x ab.

Nun kann man für Vektorraumbündel wie für Geradenbündel den Begriff „negativ" definieren. Da aber Kodaira und Nakano in Zusammen-

hang mit ihrem „vanishing theorem" das Wort negativ bereits verwendet haben und unser Begriff wesentlich schwächer ist, schreiben wir:

DEFINITION 16. *Ein Vektorraumbündel V über einem kompakten komplexen Raum heißt schwach negativ, wenn es eine Umgebung $T=T(\mathfrak{O})\leqslant V$ um die Nullschnittfläche $\mathfrak{O}\subset V$ gibt, die streng pseudokonvex ist.*

Ist nun X eine beliebige komplexe Mannigfaltigkeit und $A\subset X$ eine nirgends diskrete, singularitätenfreie kompakte analytische Teilmenge von einer Kodimension $q>0$, so ist das Bündel der kontravarianten Normalvektoren $N=N(A)$ definiert. Es ist ein komplex-analytisches Vektorraumbündel über A, dessen Fasern die Dimension q haben. Man zeigt leicht: Ist N schwach negativ, dann gibt es eine streng pseudokonvexe Umgebung $T=T(A)\subset\subset X$, derart, daß A maximale kompakte analytische Teilmenge von T ist. Es folgt also (vgl. [13]):

SATZ 12. *Es sei $N(A)$ schwach negativ, dann ist A exzeptionell.*

Leider ist dieses Kriterium nur notwendig, wenn die kompakte komplexe Mannigfaltigkeit A eine einfache Struktur hat, z. B. wenn sie homogen oder wenn sie eindimensional ist. Es gibt bereits 3-dimensionale komplexe Mannigfaltigkeiten X mit 2-dimensionalen singularitätenfreien exzeptionellen analytischen Mengen A, derart, daß $N(A)$ nicht schwach negativ ist.

Ist X ein allgemeiner komplexer Raum, $A\subset X$ eine beliebige niederdimensionale, nirgends diskrete kompakte analytische Teilmenge, so besitzt A im allgemeinen kein Normalenbündel. Man kann sich jedoch auf folgende Weise helfen, wobei man wesentlich den Begriff der kohärenten analytischen Garbe, wie er von H. Cartan definiert wurde, heranzieht. Es sei $\mathfrak{m}$ eine kohärente Garbe von Keimen von holomorphen Funktionen, die auf A verschwinden. $\mathfrak{m}$ ist also eine Untergarbe der Garbe $\mathfrak{n}$ aller Keime auf A verschwindender holomorpher Funktionen und damit auch eine Untergarbe von $\mathcal{O}=\mathcal{O}(X)$. Jeder Halm $\mathfrak{m}_x$ ist ein Ideal in dem Ring $\mathcal{O}_x$. Wir setzen zusätzlich voraus, daß in $X-A$ gilt: $\mathcal{O}=\mathfrak{m}$. Unter $\mathfrak{m}\cdot\mathfrak{n}$ verstehen wir die Garbe, die in jedem Punkt $x\in X$ als Halm das Produkt der Ideale $\mathfrak{m}_x$ und $\mathfrak{n}_x$ hat. $\mathfrak{m}\cdot\mathfrak{n}$ ist also eine Untergarbe von $\mathfrak{m}$ und man hat $\mathfrak{m}\cdot\mathfrak{n}|X-A=\mathfrak{m}|X-A=\mathcal{O}|X-A$. Es folgt nun, daß $S=\mathfrak{m}/\mathfrak{m}\cdot\mathfrak{n}|A$ eine kohärente analytishe Garbe über A ist. Im Falle, daß X eine komplexe Mannigfaltigkeit und $A\subset X$ singularitätenfrei eingebettet und $\mathfrak{m}=\mathfrak{n}$ ist, kann man die Struktur von S einfach bestimmen. S ist nichts weiter als die Garbe der Keime von lokalen holomorphen Schnittflächen in dem Bündel $V=V(A)$ der kovarianten Normalenvektoren an A. Das zu V duale Vektorraumbündel ist $N(A)$. Im allgemeinen Falle kann man zu S ebenfalls ein „Duales" konstruieren. Dieses sei mit $N_{\mathfrak{m}}(A)$ bezeichnet. Es ist im allgemeinen kein Vektorraumbündel mehr, sondern nur noch ein „linearer Raum über X", der zwar als Fasern noch komplexe Vektorräume aufweist, sich aber wesentlich von einem Vektorraumbündel dadurch unterscheidet, daß die Dimension der Fasern von $x\in X$ abhängt. In dem vorhin betrachteten Sonderfall gilt jedoch $N_{\mathfrak{m}}(A)=N(A)$.

Wie ein Vektorraumbündel besitzt jeder lineare Raum über X eine Nullschnittfläche $\mathfrak{O}$. Wir nennen $N_{\mathfrak{m}}(A)$ schwach negativ, wenn es eine streng pseudokonvexe Umgebung $T=T(\mathfrak{O})\leqslant N_{\mathfrak{m}}(A)$ gibt. Man zeigt wieder: Existiert $T(\mathfrak{O})$, so gibt es auch eine streng pseudokonvexe Umge-

bung $T^* = T^*(A) \leqslant X$, so daß A maximale kompakte analytische Teilmenge von T^* ist. Es folgt:

SATZ 13. *Ist $N_{\mathfrak{m}}(A)$ schwach negativ, so ist A eine exzeptionelle analytische Teilmenge von X.*

Es darf vermutet werden, daß dieses Kriterium auch notwendig ist. Setzt man A als exzeptionell voraus, so dürfte es eine kohärente Garbe $\mathfrak{m}$ geben, die die geforderten Eigenschaften hat, so daß $N_{\mathfrak{m}}(A)$ schwach negativ ist. Der Beweis dieser Aussage ist jedoch wahrscheinlich schwierig. Wir spezialisieren uns nun wieder auf den Fall, daß X eine 2-dimensionale komplexe Mannigfaltigkeit ist. $A \subset X$ sei eine nirgends dichte, nirgends diskrete kompakte analytische Teilmenge. Sie zerfalle in die irreduziblen Komponenten $A_1, \ldots, A_k$. Nach einem Satz der Matrizentheorie ist nun die Schnittmatrix $(a_{\nu\mu})$ mit $a_{\nu\mu} = (A_\nu, A_\mu)$ genau dann negativ definit, wenn es natürliche Zahlen $c_1, \ldots, c_k > 0$ gibt, so daß $d_\nu = \sum_\mu a_{\nu\mu} c_\mu < 0$ gilt. Andererseits kann man zeigen (die eine Richtung folgt aus Satz 13), daß A genau dann exzeptionell ist, wenn es natürliche Zahlen $c_1, \ldots, c_k$ gibt, so daß die Beschränkung von F auf A negativ ist, wobei F das zu dem Divisor $\sum c_\nu A_\nu$ gehörende Geradenbündel bezeichnet. $F|A$ ist aber genau dann negativ, wenn $d_\nu = \sum_\mu a_{\nu\mu} c_\mu < 0$ für $\nu = 1, \ldots, k$ gilt. Somit haben wir bewiesen:

SATZ 14. *A ist genau dann exzeptionell, wenn die Schnittmatrix $(a_{\nu\mu})$ negativ definit ist.*

Dieses Resultat wurde in einem Spezialfall zuerst von Mumford hergeleitet. Mumford setzt zusätzlich voraus, daß jede irreduzible Komponente A_ν ein singularitätenfrei eingebetteter P^1 ist, daß der Schnitt $A_\nu \cap A_\mu$ entweder leer oder transversal ist und schließlich, daß es beliebig kleine (sich anschmiegende) Umgebungen $U = U(A)$ gibt, deren Rand eine S^3 ist. Unter diesen Voraussetzungen kann er dann sogar zeigen, daß A, falls $(a_{\nu\mu})$ negativ definit ist, zu einem regulären Punkt niedergeblasen werden kann.

Es bleibt noch zu untersuchen, wann man durch Niederblasen aus kompakten, projektiv algebraischen Räumen wieder projektiv algebraische erhält. Schon Enriques zeigte folgende Aussage:

Es sei X eine kompakte, projektiv algebraische, 2-dimensionale komplexe Mannigfaltigkeit, $A \subset X$ sei ein singularitätenfrei eingebetteter P^1, dessen Selbstschnittzahl -1 ist. Dann läßt sich A zu einem regulären Punkt niederblasen. Der dadurch erhaltene Raum Y ist wieder projektiv algebraisch.

Kodaira hat diesen Satz mit Hilfe seiner Potentialtheorie auf Dimensionen >2 verallgemeinert. Wie wir jedoch in § 3 gesehen haben, darf man bei beliebigen exzeptionellen analytischen Mengen A nicht erwarten, daß durch Niederblasen von A projektiv algebraische Räume in algebraische Räume transformiert werden. Ein notwendiges und hinreichendes Kriterium dürfte sich nur schwer angeben lassen, so daß man in der algebraischen Geometrie einfache Aussagen über exzeptionelle analytische Mengen nicht erwarten darf.

Wir geben hier ein hinreichendes Kriterium. Es bezeichne X eine komplexe Mannigfaltigkeit, $A \subset X$ eine kompakte, singularitätenfrei eingebettete analytische Menge. Die Kodimension von A sei gleich eins. Das Normalenbündel $N = N(A)$ ist dann ein komplex-analytisches Geradenbündel über A. Wir setzen es als negativ voraus. A ist also exzeptionell und, da $F = N^*$

ein reiches Geradenbündel über A ist, auch eine kompakte projektiv algebraische Mannigfaltigkeit.

Wie üblich, bezeichnen wir mit $H^1(A, \underline{N}^s)$ die erste Kohomologiegruppe über A mit Koeffizienten in der Garbe der Keime von lokalen holomorphen Schnittflächen in N^s (für $s=1,2,3,\ldots$). Da A eine Kählersche Mannigfaltigkeit ist, verschwindet $H^1(A, \mathcal{O}(A))$ genau dann, wenn die erste Bettische Zahl von A verschwindet. Wir schreiben $H^1(A, \mathcal{O}(A)) = H^1(A, \underline{N}^0)$. Unter „Gruppe von Néron-Severi" wollen wir hier den Untervektorraum $S \subset H^2(A, R)$ verstehen, der von den Chernschen Klassen von komplex-analytischen Geradenbündeln über A aufgespannt wird. Da A als projektiv-algebraisch vorausgesetzt ist, muß S mindestens eindimensional sein. Bezeichnet $H^{1,1}(A)$ den Vektorraum der reellen harmonischen Differentialformen $\varphi = i \sum g_{\nu\mu} dz_\nu \wedge d\bar{z}_\mu$ bezüglich einer Kählerschen Metrik, so ist $H^{1,1}(A)$ in natürlicher Weise in $H^2(A, R)$ eingebettet. S ist bzgl. dieser Einbettung ein Untervektorraum von $H^{1,1}(A)$.

SATZ 15. *Es sei* dim $S=1$ *und* $H^1(A, \underline{N}^s)=0$, $s=0,1,2,\ldots$ *und* X *eine kompakte projektiv algebraische Mannigfaltigkeit. Dann ist der kompakte normale komplexe Raum, den man durch Niederblasen von* A *erhält, wieder projektiv algebraisch.*

Die Voraussetzungen sind erfüllt, wenn $A \simeq P^{n-1}$ und der Selbstschnitt von A homolog zu $-kH$, zu einem negativen Vielfachen einer Hyperebene in A ist (andere Möglichkeiten sind nur, daß der Selbstschnitt homolog zu einem positiven Vielfachen von H oder nullhomolog ist. A ist genau in dem ersten Fall exzeptionell). Man kann nun die Struktur der Singularität, die man durch Niederblasen von A erhält, errechnen. Im Falle $k=1$ kommt dabei heraus, daß man einen regulären Punkt bekommt. Damit ist der Kodairasche Satz aufs Neue mit Hilfe der Levischen Aussage und ohne Verwendung von Potentialtheorie bewiesen.

Die hier angegebenen Beispiele mögen genügen, um aufzuzeigen, wie man die verallgemeinerte Levische Aussage in der Theorie der komplexen Räume und der algebraischen Varietäten verwenden kann. Viele Sätze, die man früher auf potentialtheoretische Weise hergeleitet hat, lassen sich neu beweisen und auf komplexe Räume übertragen. Jedoch muß man die hier behandelten Methoden als ein zusätzliches Hilfsmittel ansehen, das keineswegs die Anwendung von partiellen Differentialgleichungen überflüssig macht. Einige Sätze von Kodaira über kompakte Kählersche Mannigfaltigkeiten, wie z. B. die Zerlegungssätze und das „vanishing theorem", lassen sich anscheinend nicht mit Hilfe des Levischen Problems behandeln. Es wäre daher von besonderem Wert, Differentialgleichungsmethoden zu gewinnen, die man auch in singulären Punkten verwenden kann.

Abschließend sei noch erwähnt, daß das Levische Problem neuerdings auch einen Einfluß auf die Theorie der partiellen Differentialgleichungen ausübt. Ehrenpreis, Hörmander, Malgrange haben gezeigt, daß in sehr allgemeinen Fällen Lösungen partieller Differentialgleichungen ein ähnliches Verhalten an Rändern aufweisen können wie holomorphe Funktionen.

Anmerkung. Ein geringer Teil der Sätze und Definitionen dieses Aufsatzes ist bereits in [12] gebracht worden. Diese Resultate mußten hier der Verständlichkeit und der Vollständigkeit halber wiederholt werden.

LITERATUR

[1]. ANDREOTTI, A. & GRAUERT, H., Théorèmes de finitude pour la cohomologie des espaces complexes. *Bull. Soc. Math. France* (1962).

[2]. BEHNKE, H., Natürliche Grenzen. *Abh. math. Sem. Hamburg*, 5 (1927).

[3]. BEHNKE, H. & STEIN, K., Modifikationen komplexer Mannigfaltigkeiten und Riemannscher Gebiete. *Math. Ann.* 124 (1951), 1–16.

[4]. BISHOP, E., Mappings of partially analytic spaces. *Amer. J. Math.* 83 (1961), 209–242.

[5]. BREMERMANN, H., Über die Äquivalenz der pseudokonvexen Gebiete und der Holomorphiegebiete im Raum von n komplexen Veränderlichen. *Math. Ann.* 128 (1954), 63–91.

[6]. CARTAN, H., Quotients of complex analytic spaces. *Contributions to Function Theory*. Bombay, 1960.

[7]. FABER, G., Über zusammengehörige Konvergenzradien von Potenzreihen mehrerer Veränderlicher. *Math. Ann.*, 61 (1905).

[8]. GODEMENT, R., *Topologie Algébrique et Théorie des Faisceaux*. Hermann, Paris, 1958.

[9]. GRAUERT, H. & REMMERT, R., Zur Theorie der Modifikationen. I. Stetige und eigentliche Modifikationen komplexer Räume. *Math. Ann.*, 129 (1955), 274–296.

[10]. —— Komplexe Räume. *Math. Ann.*, 136 (1958), 245–318.

[11]. GRAUERT, H., On Levi's problem and the imbedding of real-analytic manifolds. *Ann. Math.*, (2) 68 (1958), 460–472.

[12]. —— Die Riemannschen Flächen der Funktionentheorie mehrer Veränderlichen. *Proc. Internat. Congress Math. Edinburg*, 1958.

[13]. —— Über Modifikationen und exzeptionelle analytische Mengen. *Math. Ann.*, 146 (1962), 331–368.

[14]. GROTHENDIECK, A., Eléments de géométrie algébrique. II. *Publ. Math. (I.H.E.S.), Paris*, 8 (1961).

[15]. KODAIRA, K., On Kähler varieties of the restricted type. *Ann. Math.*, (2) 60 (1954), 28–48.

[16]. KRZOSKA, J., *Über die natürlichen Grenzen analytischer Funktionen mehrer Veränderlicher*. Dissertation, Greifswald, 1933.

[17]. LEVI, E. E., Studii sui punti singolari essenziali delle funzioni anal. di due o più var. compl. *Ann. Mat. Pura Appl.*, (3) 17 (1910).

[18]. NARASIMHAN, R., The Levi problem for complex spaces. *Math. Ann.*, 142 (1961), 355–365.

[19]. NORGUET, F., Sur les domaines d'holomorphie des fonctions uniformes de plusieurs variables complexes. *Bull. Soc. Math. France*, 82 (1954), 137–159.

[20]. OKA, K., Sur les fonctions analytiques de plusieurs variables. VI. Domaines pseudoconvexes. *Tôhoku Math. J.*, 49 (1942) 19–52.

[21]. —— Sur les fonctions analytiques de plusieurs variables. IX. Domaines finis sans point critique intérieur. *Japan J. Math.*, 23 (1954), 97–155.

[22]. REMMERT, R., Sur les espaces analytiques holomorphiquement séparables et holomorphiquement convexes. *C.R. Acad. Sci. Paris*, 243 (1956), 118–121.

[23]. ROSSI, H., Analytic spaces with compact subvarieties. *Math. Ann.*, 146 (1962), 124–145.

[24]. SCHWARTZ, L., Homomorphismes et applications complètement continues. *C.R. Acad. Sci. Paris*, 236 (1953), 2472–2475.

PROBLEMS OF STABILITY AND ERROR PROPAGATION IN THE NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

By PETER HENRICI

The organizing committee of this Congress has invited me to give a survey of a branch of numerical analysis which I consider important. The name "numerical analysis" is of recent origin; it first became part of the common mathematical usage when the Institute for Numerical Analysis was founded in Los Angeles about 14 years ago. Since then, the word has become associated with such diverse activities as airline scheduling, character recognition, and the automatic printing of telephone bills. Among mathematicians, however, the term ought to be used in a more restricted sense. What I understand by numerical analysis could more accurately be described as the study of constructive methods, or algorithms, in analysis (and algebra). In this sense almost all analysis up to 1800, and the greater part of it up to 1900 was, of course, numerical analysis. Only in this century, due to the influence of set theory, has analysis gradually evolved into an abstract rather than computational discipline, until it finally became necessary to invent a special term for what was left of algorithmic analysis.

Even if understood in this narrow sense, numerical analysis is too broad a field to be given even a superficial survey in a talk such as this. The small segment of research in this science which will be the subject of my lecture this morning is, however, in many ways typical for research in other branches of numerical analysis as well. Moreover, the study of the numerical integration of *ordinary* differential equations has enjoyed the advantage of being supported by a fully developed body of knowledge of the theoretical aspects of differential equations, both linear and nonlinear. In contrast to this, relatively little is known, for instance, about the numerical solution of nonlinear *partial* differential equations; and in order to gain an understanding of the solution even of linear partial differential equations it frequently has been necessary to consider rather special problems such as the Dirichlet problem for a rectangle.

We shall consider the initial value problem

$$x' = f(t,x) \quad (x(a) = s), \tag{1}$$

where $' = d/dt$. Here x and s are vectors in complex Euclidian N-space K_N; f is a continuous function which, in order to avoid unessential complication, we assume to be from $I \times K_N$ to K_N, where I denotes a closed finite interval $[a,b]$. In addition we assume that f satisfies a Lipschitz condition

$$\|f(t,x) - f(t,y\| \leqslant L\|x-y\| \quad (x,y \in K_N,\ t \in I), \tag{2}$$

where $\|\cdot\|$ denotes any norm on vectors in K_N, and where L is independent of x, y, and t. In this comfortable setting we can guarantee the existence of a unique solution $x=x(t)$ of problem (1) for arbitrary $s \in K_N$. The subject of my talk is the construction of approximations to this solution.

Most numerical algorithms for approximating solutions of differential equations consist in a prescription for calculating a sequence X of vectors X_n, $n=0,1,2,\ldots$, intended to approximate $x(t_n)$, the exact solution at the point $t_n=a+nh$, where h is a positive constant, called the *step* of the integration method. We shall concentrate on one important class of methods for determining the sequence X, namely the class of linear multi-step methods, first considered in full generality by Dahlquist [2]. This class contains the majority of the methods used in practice. Let $k>0$ be an integer, and let ϱ and σ be polynomials with complex coefficients whose degrees are k and $\leqslant k$, respectively. Denoting by E the displacement operator,

$$(EX)_n = X_{n+1},$$

the linear multi-step method (ϱ,σ) consists in determining the sequence X by solving the difference equation

$$\varrho(E)X_n - h\sigma(E)f(t_n, X_n) = 0. \tag{3}$$

The Euler-Cauchy method, for instance, is defined by the polynomials $\varrho(z)=z-1$, $\sigma(z)=1$, yielding the difference equation

$$X_{n+1} - X_n - hf(t_n, X_n) = 0.$$

If f satisfies (2), and if h is sufficiently small, it is easy to show that the difference equation (3) has a unique solution X (as long as $t_n \in I$) for an arbitrary set of starting vectors $X_0, X_1, \ldots, X_{k-1}$. The method does not say, however, how to find these starting vectors. It must be supplemented by a *starting procedure* Σ. By this we mean a set of k vector-valued functions $S_0(h), S_1(h), \ldots, S_{k-1}(h)$, defined for sufficiently small $h>0$. The starting algorithm then consists in putting

$$X_\nu = S_\nu(h) \quad (\nu = 0, 1, \ldots, k-1). \tag{4}$$

In the same way as the differential equation (1) can be regarded as an operator mapping each starting vector s onto a function $x(t)$, the method (ϱ,σ) can be regarded as a (non-linear) operator mapping each starting procedure Σ onto a family of sequences X, depending on the parameter h.

The main interest in the mathematical theory of numerical integration centers around the *error* of a particular method. Actually, two kinds of error must be considered. The first is the *discretization error*,

$$E_n = X_n - x(t_n).$$

It is uniquely determined by the starting procedure Σ, the continuation method (ϱ,σ), and the step h. The other error is also a kind of discretization error, arising, however, not from discretizing the interval I, but from discretizing the number system. If the algorithm described above is carried out on a (non-idealized) computing machine, it produces, in place of vectors X_n,

whose components are defined by Dedekind cuts, certain vectors $\tilde{X}_n$ with complex components whose real and imaginary parts are rational numbers of a very special form. The difference

$$R_n = \tilde{X}_n - X_n$$

is called the *round-off error*; it obviously depends not only on the method, but also on the computing machine used, the order in which the arithmetic operations are arranged, and on other details of programming.

Let us begin by considering the discretization error. As elsewhere, the sharpness of our results varies with the strength of our hypotheses. At the one extreme there is the following question: Making no hypotheses other than those already mentioned, are there any simple necessary and sufficient conditions for *convergence* of the method (ϱ,σ)? The first problem is to make this question precise. Since the sequence X does not only depend on the method (ϱ,σ), but also on the starting procedure Σ, there has to be some restriction on Σ. All we shall require is that

$$\lim_{h\to 0+} S_\nu(h) = s \quad (\nu = 0, 1, \ldots, k-1). \tag{5}$$

If this condition is satisfied, we shall say that the starting procedure is *compatible* with the initial value problem (1). We now call the method (ϱ,σ) convergent, if and only if, for all initial values problems (1) and all compatible starting procedures,

$$\lim_{h\to 0+} \sup_{t_n \in I} \|X_n - x(t_n)\| = 0. \tag{6}$$

We thus require that the solution of the difference equation tends to the solution of the differential equation as $h \to 0+$ for all initial value problems of the form (1), provided that the starting values of the difference solution tend to the starting value of the differential solution.

It is clear that for a convergent method the difference equation (3) must be a good *local* approximation to the differential equation (1). This is called the condition of *consistency* and can be formulated precisely as follows. If x is the solution of (1), and if Y is the sequence defined by

$$Y_n = x(t_n) \quad (t_n \in I),$$

then

$$\sup_{t_n \in I} \|\varrho(E) Y_n - h\sigma(E) f(t_n, Y_n)\| = o(h) \quad (h \to 0),$$

for all initial value problems (1). It is not hard to show that the consistency condition is equivalent to the two algebraic conditions

$$\varrho(1) = 0, \quad \varrho'(1) = \sigma(1). \tag{7}$$

It was recognized by J. Todd [13] in 1950 that consistency alone is not sufficient for convergence. We can easily see why by considering a simple example. Let us integrate the scalar initial value problem

$$x' = 0, \quad 0 \leqslant t \leqslant 1 \quad (x(0) = 0)$$

by the consistent method defined by

$$\varrho(z) = -z^2 + 4z - 3, \quad \sigma(z) = 2.$$

(This method arises in a natural way by using a one-sided difference approximation for the derivative.) There results the difference equation

$$-X_{n+2} + 4X_{n+1} - 3X_n = 0$$

with the solution

$$X_n = (1/2)\,\{3S_0(h) - S_1(h) + [S_1(h) - S_0(h)]3^n\}.$$

Even if the starting method is compatible, sup $\|X_n\|$ need not tend to zero as $h \to 0$ due to the very rapidly growing term 3^n. We can thus see that

$$\varrho(z) = 0 \Rightarrow |z| \leqslant 1 \tag{8a}$$

is a further necessary condition for convergence. Dahlquist [2] in 1956 showed that the less obvious condition

$$\varrho(z) = \varrho'(z) = 0 \Rightarrow |z| < 1 \tag{8b}$$

(no multiple roots on $|z| = 1$) is also necessary for convergence and proceeded to prove the considerably deeper result that the two conditions (8) (which are called the conditions of *stability*) together with the consistency condition are also sufficient for convergence. (Dahlquist used a slightly different concept of convergence, but his argument carries over with little change to the case considered here [5].)

The above formulation of the stability condition is purely algebraic and makes sense only in connection with the methods (ϱ, σ). It would seem desirable to have on hand a more intrinsic formulation of the stability condition which could be applied to a wider class of integration methods. One way how this might be accomplished is shown by the work of Lax and Richtmyer [8], who considered abstract Cauchy problems of the form

$$\begin{cases} \dfrac{d}{dt}u(t) = Au(t), & a \leqslant t \leqslant b, \\ u(a) = u_0. \end{cases} \tag{9}$$

Here u is an element of a Banach space, and A is a *linear* operator. Lax and Richtmyer showed that a finite difference operator for calculating approximations to solutions of (9) must satisfy, in addition to an obvious condition of consistency, a condition of *uniform boundedness* (with respect to the step h) in order to be convergent. In a similar vein it can be shown that the algebraic stability condition considered above is equivalent to the following: Given any initial value problem (1) and any *bounded* (but not necessarily compatible) starting procedure Σ, the resulting family of sequences X is always uniformly bounded as $h \to 0$ [6].

After discussing convergence in the most general fashion possible, let us now assume that the solution $x(t)$ is a certain number of times differentiable, and that *a priori* bounds for these derivatives are known. A long sequence of papers deal with the problem of finding *a priori* bounds for the discretiza-

tion error under these circumstances. For certain special multi-step methods such bounds were constructed for the first time by von Mises [16] in 1929. We explain the principle by considering the Euler–Cauchy method

$$X_{n+1} = X_n + hf(t_n, X_n). \tag{10}$$

We subtract from (10) the corresponding relation satisfied by the exact solution,

$$x(t_{n+1}) = x(t_n) + hf(t_n, x(t_n)) + \frac{h^2}{2} d_{n+1},$$

where the last term represents the local discretization error, which is small by virtue of consistency. It can be shown that

$$\|d_{n+1}\| \leqslant \|x''\|, \quad \text{where} \quad \|x''\| = \sup_{t \in I} \|x''(t)\|.$$

We find
$$E_{n+1} = E_n + h[f(t_n, X_n) - f(t_n, x(t_n))] - \frac{h^2}{2} d_{n+1} \tag{11}$$

and, taking norms and using the Lipschitz condition,

$$\|E_{n+1}\| \leqslant (1 + hL)\|E_n\| + \frac{h^2}{2}\|x''\|.$$

Assuming $E_0 = 0$, there readily follows by induction

$$\|E_n\| \leqslant \frac{(1+hL)^n - 1}{L} \frac{h}{2} \|x''\|$$

or, using Bernoulli's inequality and giving away only little when h is small,

$$\|E_n\| \leqslant \frac{e^{(t_n - a)L} - 1}{L} \frac{h}{2} \|x''\|. \tag{12}$$

The above bound is typical for the bounds obtained by von Mises and many that came after him. The bound always increases exponentially, even in cases where the error is known to decrease, as for instance in the numerical integration of differential equations of the form $x' = Ax$, where the eigenvalues of the matrix A have negative real parts. The von Mises bound thus does not correctly reflect the true behavior of the error for large values of t. In this respect some progress can be reported, primarily due to the work of Dahlquist [3] and Lozinskii [9]. We shall continue to use the Euler-Cauchy method as our vehicle of illustration. The trouble with the von Mises bound is that we have been too impatient to take norms. Let us return once more to the error equation (11). Denoting by $J = f_x$ the Jacobian matrix of the function f, it is readily shown that

$$f(t_n, X_n) - f(t_n, x(t_n)) = \hat{J}_n E_n,$$

where
$$\hat{J}_n = \int_0^1 J(t_n, x(t_n) + \lambda E_n)\, d\lambda.$$

We thus can write (11) in the form

$$E_{n+1} = E_n + h\hat{J}_n E_n - \frac{h^2}{2} d_{n+1}.$$

The solution of this difference equation can be represented in the form

$$E_n = -\frac{h^2}{2} \sum_{k=1}^{n} \left[\prod_{m=0}^{n-k} (I + h\hat{J}_{n-m}) \right] d_k \tag{13}$$

(I = identity matrix). We now define for $h > 0$

$$\mu(J, h) = \sup_{(t,x)} \frac{\|I + hJ(t, x)\| - 1}{h}$$

and obtain from (13), taking norms,

$$\|E_n\| \leqslant \frac{h^2}{2} \|x''\| \sum_{k=1}^{n} (1 + h\mu(J, h))^{n-k},$$

$$\leqslant \frac{h}{2} \|x''\| \frac{e^{(t_{n-a})\mu(J,h)} - 1}{\mu(J, h)}.$$

This is a bound of the same type as (12), with the exception that the Lipschitz constant is replaced by $\mu(J,h)$. The point is that $\mu(J,h)$ is usually smaller than the Lipschitz constant and may even be negative. For the sake of illustration, let us consider a system of equations with constant coefficients, $x' = Ax$. Here $J = A$, $L = \|A\|$. On the other hand

$$\mu(A) = \lim_{h \to 0+} \mu(A, h) = \lim_{h \to 0+} \frac{\|I + hA\| - 1}{h}$$

exists and is always $\leqslant \|A\|$. If $\|\cdot\|$ denotes the spectral norm, then

$$\mu(A) = \text{largest eigenvalue of } \tfrac{1}{2}(A + A^*).$$

The quantity $\mu(A)$ has been called the *logarithmic norm* of the operator A by Lozinskii [9], who used it to obtain non-exponential error bounds for Adams' methods. Working independently, Dahlquist [3] has used $\mu(A)$ to obtain similar bounds for another class of linear multi-step methods. (The intersection of the two classes is Euler's method.) Some earlier results of Uhlmann [14] go in the same direction.

In the above discussion we have tried to convey the idea that an error bound should at the same time approximate the actual behavior of the error in order to be useful. From this point of view the most satisfactory statement about the error is an asymptotic formula. Ideally, such a formula takes the form

$$E_n = h^p e(t_n) + 0(h^{p+1}). \tag{14}$$

Here p is a known integer, and $e(t)$ is a known function that does not depend on h. If an explicit bound for the term $0(h^{p+1})$ is given, (14) also furnishes

a bound for the error, and this bound is best possible in the sense that the coefficient of the leading power of h cannot be improved. (Any proof that $h^p e(t_n)$ is asymptotic to the error must contain some sort of estimate of the remainder. How explicit this estimate is made is mainly a question of esthetics of the presentation of mathematics.) There are other reasons why one might be interested in an asymptotic error formula of type (14), or at least in the fact that such a formula exists. The asymptotic formula is the only means to compare objectively the accuracy of different methods of integration. (From the fact that an error *bound* is better for one method than for another, it does not follow that the method itself is better.) Finally, the mere fact that an asymptotic formula holds justifies the application of a convergence-accelerating procedure known as Richardson's deferred approach to the limit [11].

Granting sufficient smoothness of all functions involved, an asymptotic formula of type (14) holds for all convergent methods (ϱ,σ) provided that the starting procedure is sufficiently accurate. The integer p is equal to the *order* of the method, which may be defined as the order of the zero(1) of the analytic function $(\log z)^{-1}\varrho(z)-\sigma(z)$ at $z=1$, and the function $e(t)$ is the solution of the initial value problem

$$e'=J(t)e+Cx^{(p+1)}(t) \quad (e(a)=0),$$

where $J(t)=J(t,x(t))$, and where C is a constant depending only on ϱ and σ.

If the method is such that $z=1$ is the only zero of ϱ on the unit circle, the conditions to be imposed on Σ in order for a formula of type (14) to hold can be relaxed considerably. All we need is that $S_\nu(h)=s+O(h)$. If the differential equation is linear, the starting procedure does not even have to be compatible [6]. Some work in progress indicates that, for the methods under discussion, (14) merely represents the first term of an asymptotic expansion of the error in powers of h.

The situation is totally different if the polynomial ϱ has several zeros on the unit circle. It can be shown that no asymptotic expansion of the type just mentioned can exist in this case. In fact, if the starting values are not very accurate, not even the leading term of the error formula is as simple as indicated above. If we denote the zeros of modulus 1 of ϱ by $z_1=1, z_2, \ldots, z_m$, there have to be added to the function $e(t_n)$ in (14) terms like $z_\nu^n\, e_\nu(t_n)$, where

$$e_\nu'=\lambda_\nu J(t)e_\nu, \quad e_\nu(a)\neq 0 \quad (\nu=1,2,\ldots,m),$$

and where the λ_ν are constants that again depend only on the method,

$$\lambda_\nu=\frac{\sigma(z_\nu)}{z_\nu\varrho'(z_\nu)} \quad (\nu=1,2,\ldots,m).$$

The starting values $e_\nu(a)$ depend on the starting procedure Σ.

Dahlquist [3], who encountered the numbers λ_ν in connection with his error bounds, called them *growth parameters*, because they determine the growth of the parasitic error components e_ν. Dahlquist also proved that, if, for a given even integer $k>0$, ϱ and σ are selected such that the order p is

(1) Consistency implies that $p\geqslant 1$.

maximized without violating the condition of stability, then *all* zeros of ϱ satisfy $|z|=1$, and some of the corresponding growth parameters are always negative. Thus some of the spurious error components e_ν always grow in the wrong direction. This can be dangerous if the eigenvalues of $J(t)$ have negative real parts. The presence of such spurious error components was discussed already by Rutishauser [12] in the case of linear differential equations with constant coefficients. It, too, was called numerical instability by him and by others. However, since the methods involved are still stable according to our earlier definition, it would perhaps be better to speak of *weak stability* in their connection. The methods where $z=1$ is the only zero on the unit circle accordingly would be called *strongly stable*.

With the discussion of weak stability I have reached the end of what I wish to say about discretization error. Weak stability and growth parameters, however, play a role also in the propagation of round-off error. As already mentioned, the sequence of vectors $\tilde{X}_n$ actually calculated is not the exact solution of the basic recurrence relation (3). Instead we have

$$\varrho(E)\tilde{X}_n - h\,\sigma(E)f(t_n, \tilde{X}_n) = \varepsilon_n, \tag{15}$$

where the vectors ε_n, called the *local round-off errors*, are different from zero. By analyzing the round-off operations of the computing machine it is usually not difficult to obtain rigorous bounds for $\|\varepsilon_n\|$. By using techniques similar to those applied in the discussion of discretization error, one then may obtain rigorous bounds for the accumulated round-off error $R_n = \tilde{X}_n - X_n$. Bounds obtained in this manner are necessarily valid under the assumption that the local round-off errors have always their maximum value and reenforce each other systematically. They fail to take into account the statistical nature of the rounding process and thus overshoot the mark by a wide margin. And again, as in the case of discretization error, bounds do not necessarily yield a fair comparison between different methods and do not tell the user what kind of error he actually has to expect. For these reasons a statistical theory of round-off error, coupled with asymptotic techniques as $h \to 0$, seems to hold more promise than a completely deterministic theory. (This is not necessarily the case in other branches of numerical mathematics. Thanks to the ingenious and exhaustive researches of J. H. Wilkinson [17], fully deterministic and yet realistic round-off error bounds are available today for many algorithms of numerical linear algebra.)

Statistical considerations were applied to the numerical integration of ordinary differential equations for the first time by the eminent astronomer Dirk Brouwer [1] in 1937, who treated integration like forming a sum of numbers with random errors. For systems of two equations, and for a special integration method, a more accurate theory was proposed by Rademacher [10] in 1948, who took into account the propagation of the effect of each local error through the subsequent computation. A complete theory for arbitrary systems of equations and arbitrary linear multi-step methods has been given only recently [5, 6]. We assume that the local round-off errors are random vectors whose expected value and whose covariance matrix are known and are of the form.

$$E(\varepsilon_n) = \mu q(t_n) \qquad \text{covar}\,(\varepsilon_n) = \sigma^2 C(t_n),$$

where μ and σ are $o(h)$, but independent of t, and where the functions q and C satisfy certain mild regularity conditions. (For operations in fixed point arithmetic it is frequently permissible to take $q=0$, $C=I$, for floating point operations, $q=x$ frequently leads to realistic results.) We also assume (with Rademacher and Brouwer) that all components of the vectors ε_m and ε_n are *independent for* $m \neq n$. Under these hypotheses the problem of determining the parameters of the distribution of the accumulated round-off error becomes tractable, essentially by perturbation methods. We find that R_n has the expected value

$$E(R_n) = \frac{\mu}{h\varrho'(1)}\{m(t_n) + O(h \log h)\},$$

where
$$m' = J(t)m + q(t) \quad (m(a)=0).$$

The method enters the picture only through the factor $1/\varrho'(1)$; weak stability does not play a role here. This is not the case, however, for the covariance matrix. Here we find

$$\operatorname{covar}(R_n) = \frac{\sigma^2}{h}\{V(t_n) + O(h \log h)\}.$$

where
$$V(t) = \sum_{\nu=1}^{m} \frac{V_\nu(t)}{|\varrho'(z_\nu)|^2},$$

and the matrices V_ν are the solutions of the initial value problems

$$V_\nu' = \lambda_\nu J(t) V_\nu + \bar{\lambda}_\nu V_\nu J_\nu^*(t) + C(t) \qquad (V_\nu(a)=0).$$

The sum is with respect to all zeros $z_1, z_2, \ldots, z_m$ of ϱ having absolute value 1; λ_ν is the growth parameter corresponding to z_ν. It should be noted that the above results depend only on the method (ϱ, σ) and not on the starting procedure Σ.

It is interesting to observe the results of the above theory in idealized special cases. For the sake of illustration we shall consider two systems of four equations each, both describing the motion of a point in a (ξ, η)-plane. The first is the first order system corresponding to the two equations.

$$\xi'' = -\xi, \quad \eta'' = -\eta \quad (\textit{linear model problem}).$$

The second is the nonlinear system describing the relative position of two bodies subject to Newtonian gravitation (the *two body problem*),

$$\xi'' = -\frac{\xi}{R^3}, \quad \eta'' = -\frac{\eta}{R^3} \quad (\textit{nonlinear model problem}),$$

where $R = \sqrt{\xi^2 + \eta^2}$. Under the initial conditions

$$\xi(0)=1, \quad \xi'(0)=0, \quad \eta(0)=0, \quad \eta'(0)=1,$$

both systems have the identical solution

$$\xi(t) = \cos t, \quad \eta(t) = \sin t.$$

Yet the pattern of error propagation in the two cases is quite different, and is different also for different methods of integration [6]. Assuming that the expected value of the local round-off error is proportional to the solution, we find (denoting components by subscripts)

$$\sqrt{m_1^2(t)+m_3^2(t)}=\begin{cases} t & \text{for the linear model problem,} \\ 3t^2 & \text{for the nonlinear model problem.}\end{cases}$$

(leading terms only are given). In the table below we give the sum of the first and third diagonal elements of the covariance matrix $V(t)$ for both model problems and for three typical integration formulas.

Method	Stability	Growth parameters	Model problem: Linear	Model problem: Nonlinear
Adams	strong	1	$2t$	$2t^3$
Nyström	weak	$1, -1$	t	t^3
Milne	weak	$1, -\frac{1}{3}$	t	$1.95\, e^{0.44t}$

For all methods the error grows much more strongly in the nonlinear model problem. It is not true, however, that a more negative growth parameter implies a stronger error growth. The most astonishing feature in the above table is the exponential growth of the error in the Milne method, which completely invalidates the numerical results after, say, 10 revolutions.

Statistical results of the above type can be tested experimentally by integrating the same differential equation a large number of times under slightly varying initial conditions and comparing the results with multiple precision results [5]. In this manner an experimental distribution of round-off errors is realized, whose expected value and covariance matrix can be compared with those of the predicted distribution. Such experiments were performed for the two model problems described above; the nonlinear and exponential growth of error in the nonlinear model problem was reproduced faithfully.

This concludes what I wish to say about the numerical integration of first order systems by linear multi-step methods. In the remaining time I would like to report briefly about some extensions of the above theories.

In an important paper [15] Urabe considers integration algorithms where to the expression $\sigma(E)f(t_n, X_n)$ is added a term of the form

$$f(t_{n+k+1}, \hat{\varrho}(E)X_n - h\hat{\sigma}(E)\, f(t_n, X_n))$$

where $\hat{\varrho}$ and $\hat{\sigma}$ are two new polynomials of degree $\leqslant k$. These methods no longer are of the linear multi-step type, but rather represent a crossbreed between the latter and the methods of Runge-Kutta type. Urabe has formulated necessary and sufficient conditions for the convergence of these hybrid methods and has found bounds for the discretization error, using a generalization of Lozinskii's logarithmic norm. For judicious choices of the polynomials ϱ, σ, $\hat{\varrho}$ and $\hat{\sigma}$, these bounds are superior to the bounds that have been obtained for some conventional methods.

Some of the results for linear multi-step methods have been extended to cover methods that integrate equations of the special form

$$x^{(m)} = f(t, x) \quad (m = 2, 3, \ldots) \tag{16}$$

directly, without reduction to a first order system. The approximating difference equation is now of the form

$$\varrho(E) X_n - h^m \sigma(E) f(t_n, X_n) = 0, \tag{17}$$

which is the same as before, except that h is replaced by h^m. Dahlquist [3] has extended to this case his results on stability. For consistency, ϱ now must have a zero of order m at $z=1$, and $\varrho^{(m)}(1) = m!\, \sigma(1)$. For stability, ϱ again may have no zeros outside the unit circle, and the zeros on the unit circle must be at most of order m.

The case $m=2$ of (16) has been studied in somewhat greater detail because of its importance in celestial mechanics. Of particular interest are some results concerning propagation of round-off error [4, 5]. Making the same hypotheses as before, we now find that

$$E(R_n) = \frac{\mu}{h^2} \{ m(t_n) + O(h \log h) \}.$$

$$\text{covar}\,(R_n) = \frac{\sigma^2}{h^3} \{ V(t_n) + O(h \log h) \},$$

where m and V are known functions. This means that both means and standard deviations are worse, by a factor h^{-1}, than in the integration of first order systems. This unfavorable behavior can be corrected, however, by writing equation (17) (with $m=2$) in the "summed" form

$$\left.\begin{aligned} \frac{E\varrho(E)}{E-1} X_n - h\sigma(E)\, F(t_n, X_n) &= 0, \\ (E-1)\, F(t_n, X_n) &= hf(t_{n+1}, X_{n+1}). \end{aligned}\right\} \tag{18}$$

The mathematical solution of (18), if correctly started, is identical with that of (17); however, the round-off error distribution of the solution of (18) now resembles the distribution of the error in the first order case. A quantitative theory has been given; its results agree closely with those of numerical experiments.

The crucial hypothesis for the above statistical theories is the hypothesis of independence of local errors. While this assumption seems to yield realistic results in many cases, some situations are known, both in the numerical integration of differential equations [7] and in other iterative processes of numerical analysis, where the local errors definitely cannot be considered to be independent. To elucidate the conditions under which the local errors act like independent variables would seem to be a fascinating if difficult problem.

REFERENCES

[1]. BROUWER, D., On the accumulation of errors in numerical integration. *Astronomical J.*, 46 (1937), 149–153.

[2]. DAHLQUIST, G., Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.*, 4 (1956), 33–53.

[3]. —— Stability and error bounds in the numerical integration of ordinary differential equations. *Trans. Roy. Inst. Techn. Stockholm*, No. 130 (1959).

[4]. HENRICI, P., The propagation of round-off error in the numerical solution of initial value problems involving ordinary differential equations of the second order. *Proc. of Symposium on Numerical Treatment of Ordinary Differential Equations, Integral Equations, and Integro-Differential Equations*, 275–291. Birkhäuser, Basel, 1960.

[5]. —— *Discrete Variable Methods in Ordinary Differential Equations*. Wiley, New York, 1962.

[6]. —— *Errorpropagation for Difference Methods*. Wiley, New York, 1963.

[7]. HUSKEY, H. D., On the precision of a certain procedure of numerical integration. With an appendix by Douglas R. Hartree. *J, Research Nat. Bur. Standards*, 42 (1949), 57–62.

[8]. LAX, P. D. & RICHTMYER, R. D., Survey of the stability of linear finite difference equations. *Comm. Pure Appl. Math.*, 9 (1956), 267–293.

[9]. LOZINSKIĬ, S. M., Error estimate for numerical integration of ordinary differential equations. I. *Izv. Vysš. Učebn. Zaved. Matematika*, 1958, 52–90; errata, 1959, 222.

[10]. RADEMACHER, H., On the accumulation of errors in processes of integration on high-speed calculating machines. Proceedings of symposium on large-scale digital calculating machinery. *Ann. Comput. Labor. Harvard Univ.*, 16 (1948), 176–187.

[11]. RICHARDSON, L. F., The deferred approach to the limit. I—Single lattice. *Trans. Roy. Soc. London*, 226 (1927), 299–349.

[12]. RUTISHAUSER, H., Über die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen. *Z. Angew. Math. Physik*, 3 (1952), 65–74.

[13]. TODD, J., Notes on numerical analysis. I. Solution of differential equations by recurrence relations. *Math. Tables Aids Comput.*, 4 (1950), 39–44.

[14]. UHLMANN, W., Fehlerabschätzungen beim Anfangswertproblem gewöhnlicher Differentialgleichungssysteme 1. Ordnung. *Z. Angew. Math. Mech.*, 18 (1957), 83–90.

[15]. URABE, M., Theory of errors in numerical integration of ordinary differential equations. *J. Sci. Hiroshima Univ., Ser. A-1*, 25 (1961), 3–62.

[16]. VON MISES, R., Zur numerischen Integration von Differentialgleichungen. *Z. Angew. Math. Mech.*, 10 (1930), 81–92.

[17]. WILKINSON, J. H., Error analysis of direct methods of matrix inversion. *J. Assoc. Comput. Mach.*, 8 (1961), 281–330.

TRANSFORMÉES DE FOURIER DES FONCTIONS SOMMABLES

Par J.-P. KAHANE

Dans l'analyse de Fourier, il est sans doute peu de questions plus naturelles que l'étude des transformées de Fourier des fonctions sommables. Cela comprend, en particulier, les fonctions, continues sur le cercle, dont la série de Fourier est absolument convergente (transformées de Fourier des fonctions sommables sur le groupe des entiers), et les suites de coefficients de Fourier-Lebesgue (transformées de Fourier des fonctions sommables sur le cercle). Le sujet, quoique travaillé depuis longtemps, est resté assez mystérieux : en 1958, dans l'introduction à la seconde édition de son traité, Zygmund comptait, au nombre des deux ou trois problèmes majeurs concernant les séries trigonométriques, celui de préciser la structure des fonctions sommes de séries de Fourier absolument convergentes.

Depuis quelques années, et particulièrement depuis 1958, d'importants progrès ont été faits : des problèmes, cherchés depuis longtemps, ont été résolus; d'autres se sont posés. Je voudrais tenter de donner ici un tableau d'ensemble, en souhaitant que de nombreux congressistes aient pu prendre connaissance directement, en entendant leurs auteurs, des résultats récents les plus remarquables (je pense, en particulier, à P. Malliavin et P. J. Cohen). Les mathématiciens intéressés pourront d'ailleurs se reporter à l'excellent ouvrage de W. Rudin [69], qui va paraître incessamment.

1. Notations et introduction

Etant donné deux groupes abéliens localement compacts en dualité, G et Γ, on sait que la transformation de Fourier fait correspondre, à toute fonction sommable sur G, une fonction continue sur Γ, et nulle à l'infini si Γ n'est pas compact (sauf précision contraire, fonction = fonction à valeurs complexes). On désigne par $A(\Gamma)$ l'ensemble des fonctions continues sur Γ, transformées de Fourier de fonctions sommables sur G. Ainsi

$$f \in A(\Gamma) \Leftrightarrow f(x) = \int_G (x, u)\, \hat{f}(u)\, dm(u),$$

dm étant une mesure de Haar sur G, $\hat{f} \in L^1(G)$, et (x,u) étant la valeur en u du caractère x.

Les cas les plus classiques sont les suivants.

$G = \Gamma = R$, droite numérique :

$$f \in A(R) \Leftrightarrow f(x) = \int_R e^{2\pi i u x}\, \hat{f}(u)\, du \quad (\hat{f} \in L^1(R)).$$

$G = Z$, groupe des entiers, et $\Gamma = T$, groupe des réels modulo 1 (ou « cercle »):

$$f \in A(T) \Leftrightarrow f(x) = \sum_{n \in Z} \hat{f}(n)\, e^{2\pi i n x} \quad (\sum_{n \in Z} |f(n)| < \infty).$$

$G = T$, $\Gamma = Z$,

$$f \in \mathrm{A}(Z) \Leftrightarrow f(n) = \int_T e^{2\pi i n u} \hat{f}(u) du \quad (\hat{f} \in L^1(T)).$$

$A(\Gamma)$ est une algèbre de Banach pour la multiplication ordinaire, la norme étant $\|f\| = \int_G |\hat{f}(u)|\, dm(u)$. A chaque $x \in \Gamma$ correspond l'idéal maximal I_x des $f \in A(\Gamma)$ qui s'annulent en x, et réciproquement, tout idéal maximal s'obtient ainsi (Wiener). La topologie de Γ étant la moins fine rendant continues toute les $f \in A(\Gamma)$, on peut identifier Γ et l'espace des idéaux maximaux de $A(\Gamma)$. Etant donné un fermé $E \in \Gamma$, on désignera par I_E l'idéal des $f \in A(\Gamma)$ qui s'annulent sur E, et par $A(E)$ l'algèbre quotient $A(\Gamma)/I_E$. On peut considérer $A(E)$ comme l'algèbre des restrictions à E des $f \in A(\Gamma)$, avec la norme $\|g\|_E = \inf \|f\|$, la borne inférieure étant prise pour toutes les fonctions $f \in A(\Gamma)$ égales à g sur E.

Cette structure d'algèbre de Banach suggère des problèmes. Par exemple, quels sont les endomorphismes de $A(\Gamma)$? Existe-t-il un « calcul symbolique » dans $A(\Gamma)$ plus étendu que le calcul symbolique classique, à l'aide des fonctions analytiques ? Y-a-t-il d'autres idéaux fermés dans $A(\Gamma)$ que les I_E ? Comme nous le verrons, des progrès décisifs ont été faits ces dernières années sur ces trois problèmes.

On peut aussi—c'est le point de vue de l'analyse classique, et c'est celui que nous adoptons au début de cet exposé—considérer $A(\Gamma)$ comme une classe de fonctions dont il est intéressant d'étudier et de décrire les propriétés locales. En effet, l'appartenance à $A(\Gamma)$ est une propriété locale dans le sens suivant (Wiener) : soit f une fonction définie sur Γ; si, au voisinage de tout $x \in \Gamma$ (et, le cas échéant, du point x à l'infini), f est égale à une $f_x \in A(\Gamma)$, il s'ensuit que $f \in A(\Gamma)$. Si, par exemple, Γ est discret, on peut se demander quelles sont les majorations à l'infini qui entraînent l'appartenance à $A(\Gamma)$. Si $\Gamma = T$, on peut se demander quelles majorations du module de continuité $\omega_f(\delta) = \sup_{|x-x'| \leqslant \delta} |f(x) - f(x')|$ entraînent $f \in A(\Gamma)$. L'étude de ce dernier problème remonte à S. Bernstein [2, 3].

Dans cet esprit descriptif, on peut aussi s'intéresser aux classes $A(E)$ que nous venons de définir, et qui sont constituées des fonctions, définies sur un fermé $E \in \Gamma$, et restrictions des $f \in A(\Gamma)$. Une circonstance heureuse est que, si E est un ensemble d'entiers, on obtient la même classe $A(E)$ en considérant les restrictions à E des $f \in A(Z)$, ou les restrictions à E (plongé dans R) des $f \in A(R)$. De même, si l'on représente T par un intervalle de longueur 1 sur R, dont on identifie les extrémités, et si E est un fermé strictement intérieur à cet intervalle, les éléments de $A(E)$ peuvent être indifféremment définis à partir de $A(T)$ ou de $A(R)$. Les études dans cette direction sont un peu plus récentes, encore qu'on puisse y rattacher certains travaux de Sidon [73, 74] et de Banach [1]. Elles sont loin d'être terminées.

2. Théorie descriptive des $A(\Gamma)$

Le problème général de la théorie descriptive est le suivant : étant donné un fermé $E \subset \Gamma$, et un ensemble $\mathfrak{F}$ de fonctions continues sur Γ, décider si $\mathfrak{F}$ est contenu, ou non, dans $A(E)$.

Dès 1914, le problème s'est posé à S. Bernstein sous la forme suivante : à quelle condition, portant sur la fonction positive $\omega(\delta)$ $(\delta>0)$, $A(T)$ contient-elle l'ensemble $\Lambda_\omega(T)$ des fonctions f, définies sur T, dont le module de continuité satisfait $\omega_f(\delta)=0(\omega(\delta))$ $(\delta\to 0)$. Il n'avait pas obtenu de condition nécessaire et suffisante, mais son résultat de 1934 est néanmoins très précis : l'inclusion a lieu si la série $\sum n^{-\frac{1}{2}}\omega(1/n)$ converge (par exemple, $\omega(\delta)=\delta^{\frac{1}{2}+\epsilon}$) et elle n'a pas lieu si ω est assez régulière et que la série diverge (par exemple, $\omega(\delta)=\delta^{\frac{1}{2}}$). La condition nécessaire et suffisante n'a été obtenue qu'en 1953, par Stetchkin [77]: c'est la convergence de la série $\sum n^{-\frac{1}{2}}\omega^*(1/n)$, où ω^* désigne la fonction borne supérieure des modules de continuité inférieurs à ω. Naturellement, la partie originale du travail de Stetchkin consiste à montrer que, si la série diverge, il existe une $f\notin A(T)$ telle que $\omega_f(\delta)=0(\omega^*(\delta))$.

Ce problème se transcrit naturellement quand on considère un compact $E\subset T$ ou $E\subset R$ (voir introduction). On définit $\Lambda_\omega(E)$ comme l'ensemble des fonctions f, définies sur E, dont le module de continuité satisfait $\omega_f(\delta)=0(\omega(\delta))$ $(\delta\to 0)$; restreignons nous, pour simplifier, à des fonctions $\omega(\delta)$ positives et croissantes, telles que $\omega(\delta)/\delta$ soit décroissante. A quelle condition a-t-on $\Lambda_\omega(E)\subset A(E)$? Certains résultats précis ont été obtenus [27, 39]. Pour nous borner à un exemple, soit E un ensemble du type de Cantor, construit avec un rapport de dissection $\xi(0<\xi<\frac{1}{2})$ (cela veut dire que E est formé des points $a+b(\pm 1\pm\xi\pm\xi^2\pm\ldots)$, a et $b\neq 0$ étant donnés ; les ensembles triadiques de Cantor correspondent à $\xi=\frac{1}{3}$); pour l'inclusion $\Lambda_\omega(E)\subset A(E)$, il faut et suffit que la série $\sum n^{-\alpha/2}\omega(1/n)$ converge, où $\alpha=-\log\xi/\log 2$ est la dimension de Hausdorff de E.

Indiquons une question naturelle. A quelle condition, portant sur E, les inclusions $\Lambda_\omega(E)\subset A(E)$ et $\Lambda_\omega(T)\subset A(T)$ ont-elles lieu pour les mêmes fonctions ω? Il est nécessaire pour cela que E soit de mesure positive. Est-ce suffisant ?

Une remarque s'impose. Si f est définie sur T, on voit immédiatement si $f\in A$ ou $f\notin A$ en considérant le développement de Fourier de f. Si f est définie sur $E\subset T$, on peut montrer que $f\in A(E)$ en extrapolant f, sur T, par une fonction convenable. Mais si l'on veut montrer que $f\notin A(E)$, la méthode naturelle consiste à utiliser la dualité : les mesures $d\mu$ portées par E définissent des formes linéaires continues sur $A(E)$, de norme $\|d\mu\|_{A^*}=\sup_n|\hat{\mu}(n)|$ avec $\hat{\mu}(n)=\int e^{2\pi inx}d\mu(x)$; s'il existe une suite de mesures $d\mu_j$, portées par E, dont les normes dans A^* soient bornées, et telles que $\sup_j\left|\int_E f\,d\mu_j\right|=\infty$, on a bien $f\notin A(E)$.

L'analogue du problème de S. Bernstein pour $A(Z)$ est le suivant : à quelle condition, portant sur la suite positive $\{\omega_n\}_{n\in Z}$, est-il vrai que $A(Z)$ contient toutes les suites $f=\{f(n)\}_{n\in Z}$ telles que $f(n)=0(\omega_n)$ quand $n\to\pm\infty$? La solution en a été obtenue, en 1926, par Littlewood [50] : il faut et suffit que la série $\sum\omega_n^2$ converge; autrement dit, la condition est la même que si l'on remplace $A(Z)$ par $L^2(Z)$. La partie non triviale consiste à montrer que, si $\sum\omega_n^2=\infty$, il existe des phases φ_n telles que $\{\omega_n e^{i\varphi n}\}_{n\in Z}\notin A(Z)$. Elle a été remarquablement simplifiée par Paley et Zygmund [58] qui ont montré que, dans ce cas, on a presque sûrement $\{\pm\omega_n\}\notin A(Z)$ lorsqu'on choisit au hasard les signes $\pm$, avec probabilité $\frac{1}{2}$, et indépendamment les uns des autres.

Le théorème de Littlewood présente un caractère définitif. Mais un problème voisin, et difficile, mérite de retenir encore l'attention. Au lieu de

considérer, comme dans ce théorème, toutes les suites f astreintes à la condition $f(n)=0(\omega_n)$ $(n\to\pm\infty)$, considérons seulement les suites de coefficients de Fourier-Stieltjes (=coefficients de Fourier de mesures); à quelle condition, portant sur $\{\omega_n\}$, appartiennent-elles toutes à $A(Z)$? La condition $\sum\omega_n^2<\infty$ est encore évidemment suffisante, mais elle n'est pas nécessaire (on sait, par exemple, par un théorème de F. et M. Riesz, que $\omega_n=0$ pour n négatif suffit). Un théorème remarquable, précisant des résultats antérieurs de Wiener et Wintner [81] et Salem [70], a été publié en 1957 par Ivacheff-Moussatoff [25] : si la suite ω_n est paire $(\omega_n=\omega_{-n})$ et suffisamment régulière (si par exemple, pour chaque $\alpha>0$, la suite $\{n^\alpha\omega_n\}_{n=1,2,\dots}$ est monotone), la condition $\sum\omega_n^2<\infty$ est nécessaire et suffisante; si la série diverge, il existe une mesure singulière $d\mu$, portée par un ensemble de mesure nulle, telle que $\hat{\mu}(n)=0(\omega_n)$ $(n\to\pm\infty)$.

3. Théorie descriptive (suite) : les ensembles de Helson, Carleson, Sidon

Un cas particulier du problème général de la théorie descriptive va maintenant nous occuper. Désignons par $C(E)$ l'ensemble des fonctions continues sur E, et éventuellement nulles à l'infini. Pour quels ensembles E a-t-on $A(E)=C(E)$?

Une remarque intéressante fut faite par I. Segal en 1950 [72] : si Γ est infini, on a $A(\Gamma)\neq C(\Gamma)$. En voici une généralisation (et une démonstration) facile : appelons « ν-maille » tout ensemble de 2^ν points distincts de Γ, de la forme $a_0+\varepsilon_1a_1+\varepsilon_2a_2+\dots+\varepsilon_\nu a_\nu$, chaque $\varepsilon_j=0$ ou 1; si, quels que soit ν, E contient une ν-maille, $A(E)\neq C(E)$. En effet, si l'on avait $A(E)=C(E)$, les duals seraient les mêmes, et les mesures $d\mu$ portées par E auraient des normes $\|d\mu\|_{C^*}$ et $\|d\mu\|_{A^*}$ équivalentes ($\|d\mu\|=\int_E|d\mu|$, $\|d\mu\|_{A^*}=\sup_{u\in G}|\int(x,u)d\mu(x)|$). Or, il est aisé de voir que chaque ν-maille porte une mesure $d\mu$ telle que $\|d\mu\|_{A^*}\leqslant\varepsilon_\nu\|d\mu\|_{C^*}$, $\lim_{\nu\to\infty}\varepsilon_\nu=0$ (on peut, avec un peu de soin, choisir $\varepsilon_\nu=2^{-(\nu+1)/2}$; ce n'est pas nécessaire ici). Donc $A(E)\neq C(E)$.

Sur la droite ou le cercle, l'intérêt de considérer les compacts E pour lesquels $A(E)=C(E)$ a été reconnu par H. Helson [18]. Helson a montré qu'aucun de ces ensembles ne porte une mesure non nulle dont la transformée de Fourier soit nulle à l'infini. C'est pourquoi, de façon générale, on appelle ensemble de Helson toute partie compacte d'un groupe Γ telle que $A(E)=C(E)$.

Pour que E soit un ensemble de Helson, il est évidemment nécessaire que, pour toutes les mesures $d\mu$ portées par E, les normes $\|d\mu\|_{A^*}$ et $\|d\mu\|_{C^*}$ soient équivalentes, autrement dit, qu'il existe un $\delta=\delta(E)>0$ tel que $\|d\mu\|_{A^*}\geqslant\delta\|d\mu\|_{C^*}$ (car l'inégalité $\|d\mu\|_{A^*}\leqslant\|d\mu\|_{C^*}$ a toujours lieu). La condition est aussi suffisante.

Les ensembles de Helson ont des propriétés arithmétiques très particulières. Nous venons de voir qu'ils ne peuvent contenir des mailles dont le nombre de termes 2^ν soit arbitrairement grand. D'autre part, à tout ensemble de Helson correspond un entier K tel que, dans toute progression arithmétique de 2^s termes, il y ait au plus Ks points de E. Plus généralement, on peut choisir l'entier K de sorte que, quels que soient $x_1,\dots,x_n\in\Gamma$ et s entier il y ait dans E au plus Kns points de la forme $\alpha_1x_1+\alpha_2x_2+\dots+\alpha_nx_n$ (α_j entiers, $|\alpha_1|+\dots+|\alpha_n|\leqslant 2^s$) [38, 39].

Ces conditions sont satisfaites si E est un ensemble indépendant, c'est-à-dire qu'entre des points de E en nombre fini n'existe aucune relation linéaire non triviale à coefficients entiers. En fait, tout ensemble dénombrable indépendant est un ensemble de Helson. De même, l'ensemble des points 0 et 2^{-j} $(j=0,1,2,\ldots)$ sur la droite est un ensemble de Helson, ce qui montre qu'une progression arithmétique de 2^s termes peut en effet contenir s points de E.

L'existence d'ensembles de Helson non dénombrables n'est pas évidente. En particulier, W. Rudin a montré récemment [67] qu'il ne suffit pas que E soit indépendant pour avoir $A(E)=C(E)$.

Une manière de construire des ensembles de Helson non dénombrables est la suivante. Nous avons vu que toute mesure sur Γ définit une forme linéaire continue sur $A(\Gamma)$. De façon générale, on appellera « pseudomesure » toute forme linéaire continue sur $A(\Gamma)$; lorsque $\Gamma=R$ ou T, une pseudomesure n'est autre qu'une distribution dont la transformée de Fourier est bornée; en tous cas, on sait définir son support. On sait construire sur R ou T des ensembles parfaits sans vraie pseudomesure [38], c'est-à-dire ne portant aucune pseudomesure qui ne soit une mesure, et on voit aisément que de tels ensembles sont des ensembles de Helson. La réciproque n'est pas établie, et la question se pose, de savoir si tout ensemble de Helson est un ensemble sans vraie pseudomesure. Une réponse négative aurait un intérêt en relation avec le problème de la synthèse harmonique, que nous examinerons tout à l'heure.

La structure des ensembles de Helson est encore assez mal connue. En particulier, on ne sait pas si la réunion de deux ensembles de Helson est nécessairement un ensemble de Helson.

Avant Helson, L. Carleson [9] avait considéré les ensembles compacts $E\subset T$, tels que $C(E)=A^+(E)$, $A^+(E)$ étant l'ensemble des restrictions à E des sommes de séries « de Taylor » absolument convergentes $\sum_{n\geqslant 0}\hat{f}(n)e^{2\pi inx}$. Il avait montré que ces ensembles (« ensembles de Carleson ») sont caractérisés par la propriété suivante : il existe un $\delta_1=\delta_1(E)>0$ tel que, pour toutes les mesures $d\mu$ portées par E, $\sup_{n\geqslant 0}|\hat{\mu}(n)|\geqslant\delta_1\int|d\mu|$.

Il est évident que tout ensemble de Carleson est un ensemble de Helson, mais la réciproque n'a été établie que récemment, par I. Wik [82] : en effet, dès que E est un ensemble de Helson, les normes $\int|d\mu|$, $\sup_n|\hat{\mu}(n)$, $\sup_{n\geqslant 0}|\hat{\mu}(n)|$, $\limsup_{n\to\infty}|\hat{\mu}(n)|$ sont équivalentes.

Le théorème de Wik suggère la question suivante. Etant donné une suite d'entiers Λ et un compact $E\subset T$, désignons par $A_\Lambda(T)$ l'ensemble des fonctions sommes de séries absolument convergentes $\sum_{n\in\Lambda}\hat{f}(n)e^{2\pi inx}$, et par $A_\Lambda(E)$ l'ensemble des restrictions à E des $f\in A_\Lambda(T)$. Quelles sont les suites Λ telles que, pour tout ensemble de Carleson-Helson E, on ait $A_\Lambda(E)=C(E)$?

Ce problème se rattache au suivant, posé par A. Beurling dans son séminaire en 1959 [6] : à quelle condition, portant sur Λ et E, a-t-on $A_\Lambda(E)=A(E)$? Beurling a complètement résolu le problème lorsque E est un intervalle : la condition nécessaire et suffisante est alors $|E|<\Delta.(\Lambda)$, où $|E|$ est la mesure (longueur) de E, et $\Delta.(\Lambda)$ la densité inférieure de répartition, c'est à dire la borne supérieure des densités des suites uniformément réparties contenues dans Λ (une suite uniformément répartie, de densité D, étant une suite de la forme $\{\mu_n\}_{n\in z}$, telle que $\mu_n-n\,D=0(1)$ $(n\to\pm\infty)$). La conjecture de Beurling est que l'inégalité $|E|<\Delta.(\Lambda)$ est toujours une

condition nécessaire pour avoir $A_\Lambda(E)=A(E)$. Cela a été prouvé par H. Landau [46] dans le cas où E est la réunion de deux intervalles.

D'autres problèmes, concernant $A_\Lambda(E)$, ont été étudiés, sur lesquels nous ne nous attarderons pas [29, 34]. Certains d'entre eux font intervenir la densité supérieure de répartition (borne inférieure des densités des suites uniformément réparties contenant Λ). D'autres se rattachent à l'étude des séries de Fourier lacunaires et des suites de Sidon, dont nous allons maintenant dire un mot.

Les analogues des ensembles de Helson sur Z, à savoir les ensembles $E\subset Z$ tels que $A(E)=C(E)$, avaient été considérés depuis longtemps. En 1927 [73], Sidon avait montré que, si E est lacunaire au sens d'Hadamard (c'est-à-dire $E=\{\lambda_n\}_{n\in Z'}$, $\lambda_{-n}=-\lambda_n$ et $\inf_{n>0}\lambda_{n+1}/\lambda_n>1$), toute fonction f bornée dont la série de Fourier est de la forme $\sum_{n\in E}\hat{f}(n)e^{2\pi inx}$ (nous noterons alors $f\in L_E^\infty(T)$) est une fonction de la classe $A(T)$; autrement dit, avec les notations introduites, $L_E^\infty(T)=A_E(T)$. Puis, indépendamment semble-t-il, Banach [1] et Sidon [74] avaient observé que les propositions suivantes s'impliquent mutuellement :

a) $L_E^\infty(T)=A_E(T)$,

b) $A(E)=C(E)$, c'est à dire que toute fonction nulle à l'infini définie sur E y est la restriction d'une $f\in A(Z)$,

c) toute fonction bornée sur E y est la restriction d'une transformée de Fourier-Stieltjes.

On appelle naturellement de tels ensembles E « ensembles de Sidon », et la définition se transcrit immédiatement si l'on remplace Z par un groupe discret Γ. Ces ensembles ont de nouveau retenu l'attention depuis quelques années [77, 24, 65]. On sait qu'une réunion finie d'ensembles lacunaires à la Hadamard est un ensemble de Sidon, et qu'il existe des ensembles de Sidon qui ne sont pas de ce type. On connait des conditions suffisantes pour qu'un ensemble E d'entiers soit un ensemble de Sidon, faisant intervenir le nombre de façons d'écrire un entier n sous la forme $\pm n_1\pm n_2\ldots\pm n_s$ ($n_1<n_2<\ldots<n_s$, $n_j\in E$), soit $r(n,s)$: il suffit qu'existe une constante B telle que, pour tout n et tout s, $r(n,s)\leqslant B^s$ [77]. On connait aussi des conditions nécessaires, simples transcriptions de celles que nous avons données pour les ensembles de Helson [29]. Un outil intéressant a été découvert par Rudin [65], sous la forme d'une nouvelle proposition équivalente à *a*), *b*), *c*):

d) à toute suite réelle $\{\alpha_n\}_{n\in E}$, correspond une mesure $d\mu$ sur T et un $\delta>0$ tels que, pour tout $n\in E$, $\mathcal{R}(\hat{\mu}(n)e^{-i\alpha n})\geqslant\delta$.

De même que pour les ensembles de Helson, on ne sait pas si la réunion de deux ensembles de Sidon est un ensemble de Sidon.

4. Les changements de variable permis et les homomorphismes

Il est évident que, si $f(x)\in A(R)$, $f(ax+b)\in A(R)$ quelle que soit la fonction linéaire $ax+b$ ($a\neq 0$). En 1934, P. Lévy a posé le problème [49] : y-a-t-il d'autres fonctions φ que les fonctions linéaires pour lesquelles $f\in A(R)\Rightarrow f\circ\varphi\in A(R)$?

La réponse fut longue à venir. Elle est négative, comme le montre un théorème de Beurling et Helson [7]. Indépendamment, Leibenson [48], avec des restrictions tenant à sa méthode, établissait le résultat analogue pour

$A(T)$: les seules fonctions φ telles que $f \in A(T) \Rightarrow f \circ \varphi \in A(T)$ sont les fonctions linéaires à pente entière (voir aussi [26]).

Le problème concernant $A(Z)$ présentait des difficultés nouvelles, et sa solution, due à Rudin [60], est un peu moins simple. Elle s'exprime ainsi : pour que l'application φ de Z dans Z soit telle que $f \in A(Z) \Rightarrow f \circ \varphi \in A(Z)$, il faut et il suffit qu'existe un entier $q > 0$, et une suite $\{r(n)\}_{n \in Z}$ à valeurs entières $\neq 0$ et q-périodique ($r(n+q) = r(n)$ pour tout n), tels que, sauf pour un nombre fini de valeurs de n, on ait $\varphi(n+q) - \varphi(n) = r(n)$.

Cette question des changements de variables permis dans les classes $A(\Gamma)$, qui ressortit naturellement à des préoccupations d'analyse classique, présente un intérêt particulier si on la relie à la structure d'algèbre de Banach de $A(\Gamma)$. En effet, tout changement de variable φ de ce type définit, d'après le théorème du graphe fermé, un endomorphisme $f \to f \circ \varphi$ de $A(\Gamma)$. Inversement, soit $f \to ef$ un endomorphisme de $A(\Gamma)$. Adjoignons à Γ un point ∞ de sorte que la réunion $\Gamma_\infty = \Gamma \cup \{\infty\}$ soit compacte (∞ est isolé dans Γ_∞ si Γ est compact), et prolongeons les $f \in A(\Gamma)$ à Γ_∞ en posant $f(\infty) = 0$. Comme Γ est l'espace des idéaux maximaux de $A(\Gamma)$ (voir introduction), tout homomorphisme de $A(\Gamma)$ sur le corps des complexes est de la forme $f \to f(x)$, $x \in \Gamma_\infty$. Or, pour chaque x, $f \to ef(x)$ est un tel homomorphisme; il s'ensuit que, pour tout $x \in \Gamma_\infty$, $ef(x) = f(\varphi(x))$, φ étant une application convenable de Γ_∞ dans Γ_∞.

Plus généralement, si l'on considère deux groupes Γ' et Γ'', les homomorphismes de $A(\Gamma'')$ dans $A(\Gamma')$ sont tous de la forme $f \to f \circ \varphi$, où φ est une application de Γ'_∞ dans Γ''_∞. Ainsi, la détermination de tous les homomorphismes de $A(\Gamma'')$ dans $A(\Gamma')$ revient à celle des applications « permises » φ de Γ'_∞ dans Γ''_∞, « permise » signifiant que $f \in A(\Gamma'') \Rightarrow f \circ \varphi \in A(\Gamma')$.

Outre les exemples que nous venons de donner ($\Gamma' = \Gamma'' = R$, $\Gamma' = \Gamma'' = T$, $\Gamma' = \Gamma'' = Z$), des approches de ce problème général furent faites par Helson [16] et Rudin [62] dans le cas des isomorphismes, c'est à dire dans le cas ou φ est une bijection de Γ'' sur Γ'. La solution complète de ce problème est due à P. J. Cohen [12]. Nous allons tenter de donner une idée de la méthode et du résultat de P. J. Cohen, avant de revenir au cas particulier de $A(T)$, qui suggère quelques nouveaux problèmes.

L'idée fondamentale de P. J. Cohen est l'observation suivante : si Γ' et Γ'' sont discrets, et si φ est une application permise de Γ'' dans Γ', la fonction caractéristique du graphe de φ (fonction définie sur $\Gamma'' \times \Gamma'$) est la transformée de Fourier d'une mesure sur $G'' \times G'$. Le premier problème qui se pose (celui des « mesures idempotentes ») consiste à caractériser les parties E d'un groupe discret Γ, telles que $\sum_{x \in E}(x, u)$ soit la série de Fourier d'une mesure $d\mu(u)$ sur G. A la suite de Helson [17] et Rudin [63], qui avaient respectivement étudié les cas $\Gamma = Z$ et $\Gamma = Z^n$, Cohen étudie ce problème, et en donne la solution très simple que voici, déjà conjecturée par Rudin [64] : les ensembles E considérés sont les éléments de l'algèbre de Boole engendrée par les classes latérales de Γ (cosets) modulo ses sous-groupes ouverts. Dès lors, on a le moyen de résoudre le problème des homomorphismes de $A(\Gamma'')$ dans $A(\Gamma')$ lorsque Γ' et Γ'' sont discrets. Dans le cas général, on observe que, si φ est une application permise de Γ' dans Γ'' 1°) elle est continue 2°) elle reste permise lorsque l'on munit Γ' et Γ'' de la topologie discrète.

Cette méthode permet non seulement de caractériser tous les homo-

morphismes de $A(\Gamma'')$ dans $A(\Gamma')$, mais encore tous les homomorphismes de $A(\Gamma'')$ dans $B(\Gamma')$, $B(\Gamma')$ étant l'algèbre de Banach constituée par les transformées de Fourier des mesures sur G'. Nous n'indiquerons pas le résultat final, qui est un peu compliqué à énoncer, et nous nous bornerons à un intéressant corollaire indiqué par Rudin [69] : tout isomorphisme de $B(\Gamma'')$ sur $B(\Gamma')$ transporte $A(\Gamma'')$ sur $A(\Gamma')$; cela veut dire en particulier (pour $\Gamma'=\Gamma''=\Gamma$) que si l'on se donne l'ensemble $B(\Gamma)$ avec sa structure algébrique et elle seule, $A(\Gamma)$, comme partie de $B(\Gamma)$, est parfaitement déterminée.

La technique introduite par Cohen pour l'étude des mesures idempotentes est intéressante en elle même. Elle l'a conduit au résultat suivant (nous nous bornons à l'énoncer lorsque $\Gamma=Z$) [11] : il existe une suite $\{\omega_s\}$ positive tendant vers l'infini telle que, quels que soient les entiers distincts $n_1, n_2, \ldots, n_s$ on ait

$$\int_T \left| \sum_{j=1}^{s} e^{2\pi i n_j u} \right| du > \omega_s.$$

Il a été conjecturé par Littlewood qu'on peut choisir ω_s de l'ordre de $\log s$ (on ne peut certainement pas choisir une suite plus rapidement croissante, comme le montre l'exemple $n_j=j$). Cohen a montré que l'on peut prendre ω_s de l'ordre de $(\log s/\log\log s)^{1/8}$, et Davenport [13], améliorant sa technique, est allé jusqu'à $(\log s/\log\log s)^{1/4}$. La conjecture de Littlewood se trouve donc renforcée, et elle propose une recherche tentante aux analystes et aux arithméticiens.

Revenons au cas $A(T)$. La méthode de Cohen apparait alors comme très détournée, et on peut conclure plus facilement à l'aide de la remarque suivante [48] : les changements de variables permis φ sont tels que les fonctions $e^{2\pi i n\varphi(x)}$, images des fonctions $e^{2\pi i n x}$ de norme 1, forment un ensemble borné dans $A(T)$. Quitte à soustraire à φ une fonction linéaire de pente entière puis à la multiplier par 2π, on est ramené au problème que voici : quelles sont les fonctions réelles $\varphi\in A(T)$ telles que les normes $\|e^{in\varphi}\|$ soient $0(1)$ quand $n\to\pm\infty$? Le fait que seules les constantes satisfassent $\|e^{in\varphi}\|=0(1)$ [7] suggère de chercher quelles sont les suites $\{\omega_n\}$ croissant vers l'infini pour lesquelles seules les constantes, parmi les fonctions réelles $\varphi\in A(T)$, satisfont $\|e^{in\varphi}\|=0(\omega_n)$ $(n\to\infty)$. Il y a peu de résultats dans cette direction : on sait que, si φ est deux fois continûment dérivable et non constante, $\|e^{in\varphi}\|$ est exactement d'ordre $\sqrt{n}$, et que, si φ est linéaire par morceaux et non constante, $\|e^{in\varphi}\|$ est exactement d'ordre $\log n$ [26]. Il est raisonnable de penser que seules les constantes satisfont $\|e^{in\varphi}\|=o(\log n)$ $(n\to\infty)$.

L'intérêt de résultats de ce type serait d'approfondir l'étude des algèbres de transformées de Fourier de fonctions sommables avec poids. Disons un mot sur ce sujet. A une suite $\omega=\{\omega_n\}_{n\in Z}$, croissante pour $n>0$, et assez régulière $(\omega_{n+m}\leqslant\omega_n\omega_m)$, on attache suivant l'idée de Beurling [4] une algèbre normée $A(T;\ \omega)$, constituée des fonctions $f\in A(T)$ telles que $\|f\|_\omega=\sum_{n\in Z}|\hat{f}(n)|\omega_n<\infty$. Dire que $\|e^{2\pi i n\varphi}\|=0(\omega_n)$, c'est dire que, pour toute $f\in A(T;\omega)$, $f\circ\varphi\in A(T)$.

De manière générale, il serait intéressant d'étudier les homomorphismes d'algèbres $A(T;\ \omega)$; en d'autres termes, d'étudier les applications φ de T

dans T tels que $f \in A(T; \omega'')$ entraîne $f \circ \varphi \in A(T; \omega')$. Des résultats fragmentaires ont été obtenus [26, 41]. Il est raisonnable de conjecturer que, ω étant astreinte à certaines conditions de croissance, les seuls changements de variables permis dans $A(T; \omega)$ (cas $\omega' = \omega'' = \omega$) sont donnés par les fonctions linéaires à pente entière; Y. Katznelson a effectivement prouvé le résultat lorsque φ est supposée deux fois continûment dérivable [41].

5. Les fonctions qui opèrent et le calcul symbolique

Soit Φ une fonction à valeurs réelles définie sur l'intervalle $]-1, 1[$. On dira qu'elle « opère » dans $A(\Gamma)$ si, chaque fois que $f \in A(\Gamma)$ prend ses valeurs dans $]-1, 1[$, on a $\Phi \circ f \in A(\Gamma)$; on dit aussi qu'elle définit « un calcul symbolique » dans $L^1(G)$, ou, par abus de langage, dans $A(\Gamma)$. L'étude systèmatique des fonctions qui opèrent dans $A(T)$ remonte à l'article cité de P. Lévy. En fait, il résulte de la théorie de Wiener [80] et Lévy [49] que

a) si Φ est analytique sur $]-1, 1[$, Φ opère dans $A(T)$,
b) si Φ est analytique au voisinage de 0, et $\Phi(0) = 0$, Φ opère dans $A(Z)$,
c) si Φ est analytique sur $]-1, 1[$ et $\Phi(0) = 0$, Φ opère dans $A(R)$.

Ces trois cas sont typiques : les résultats valent si l'on remplace respectivement T, Z, R par un groupe compact, un groupe discret, un groupe non-compact et non-discret.

Pour chercher à étendre le résultat *a*), il était naturel de considérer, comme ci-dessus, la croissance des normes $\|e^{inf}\|$ quand $n \to \infty$, f étant une fonction réelle $\in A(T)$. Si en effet $\|e^{inf}\| = 0(\omega_n)$, il suffit que Φ appartienne à $A(T; \omega)$ pour que $\Phi \circ f \in A(T)$. C'est par cette méthode que Marcinkiewicz mit en évidence en 1940 des classes (f) et (Φ) telles que, si $f \in (f)$ et $\Phi \in (\Phi)$, $\Phi \circ f \in A(T)$ [57] : (f) était défini par une majoration des coefficients de Fourier, et (Φ) était une classe de fonctions indéfiniment dérivables (voir aussi Malliavin [54]).

Une extension plus remarquable du résultat *a*) est due à Beurling [5] (voir aussi R. P. Boas [8], M. Kinukawa [45]) : s'il existe une suite positive paire $\{\gamma_n\}_{n \in Z}$, décroissante pour $n > 0$, et sommable, telle que $|\hat{f}(n)| < \gamma_n$ pour tout n, alors $|f| \in A(T)$. Ce théorème conduisait naturellement à se demander si la fonction $\Phi(t) = |t|$ opère dans $A(T)$ ou, aussi bien, dans $A(Z)$; des réponses négatives furent apportées, en 1955 et 1957, au moyen de constructions laborieuses [26, 28].

En 1955, Rudin donnait la première condition nécessaire non-triviale pour que Φ opère dans $A(Z)$: Φ doit être lipschitzienne au voisinage de 0. En 1958, il fut d'abord prouvé qu'il existe des fonctions $f \in A(T)$ réelles (à savoir, certaines sommes de séries de Fourier très lacunaires) telles que $\|e^{inf}\|$ croisse presque aussi vite qu'une exponentielle; en conséquence, il existe des fonctions indéfiniment dérivables Φ qui n'opèrent pas dans $A(T)$ [30]. Puis, après quelques mois, Y. Katznelson inaugurait une série de résultats définitifs [40, 19, 20] : les conditions suffisantes, données ci-dessus en *a*), *b*), *c*), pour que Φ opère, sont aussi nécessaires. En gros, on peut dire que seules les fonctions analytiques (en se restreignant à un voisinage de 0 si Γ est discret) opèrent dans $A(\Gamma)$.

Donnons, dans le cas de $A(T)$, un aperçu de la méthode, avant d'indiquer les recherches qui ont suivi. Soit F une fonction 2π-périodique; notons

$\Sigma \gamma_n e^{inf}$ sa série de Fourier. Supposons que les translatées de F opèrent uniformément sur la fonction réelle $f \in A(T)$, dans le sens que les fonctions $x \to F(f(x)+t)$ appartiennent toutes à $A(T)$, et ont des normes bornées par un β indépendant de t. Les formules de Fourier

$$\gamma_n e^{inf(x)} = \frac{1}{2\pi} \int_0^{2\pi} F(f(x)+t)\, e^{-int}\, dt$$

donnent alors $|\gamma_n| \; \|e^{inf}\| \leqslant \beta$; ainsi, d'une minoration des normes $\|e^{inf}\|$, on tirera une majoration des γ_n [30].

Supposons maintenant que les translatées de F opèrent uniformément sur une « boule réelle » de $A(T)$, c'est à dire qu'il existe un $r>0$ et un $\beta>0$ tels que, pour toutes les fonctions réelles $f \in A(T)$ de normes $\leqslant r$, et pour tout t, $F(f+t) \in A(T)$ et $\|F(f+t)\| \leqslant \beta$. Le même calcul donne $|\gamma_n| \sup_{f \text{ réelle}, \|f\|<r} \|e^{inf}\| \leqslant \beta$. Or on établit que $\sup_{f \text{ réelle}, \|f\|<R} \|e^{if}\| = e^R$. On en déduit que les γ_n décroissent exponentiellement, donc que F est analytique [40].

Supposons enfin qu'une fonction Φ, définie sur $]-1,1[$, opère dans $A(T)$. Etant donnée une boule réelle $\mathcal{B}(0,r) : \|f\| \leqslant r < 1$, Φ définit une application $f \to \Phi \circ f$ de $\mathcal{B}(0,r)$ dans $A(T)$, que nous désignerons par (Φ). On démontre, soit en utilisant un théorème de Baire [40], soit directement [19], que Φ est un opérateur borné dans une boule $\mathcal{B}(0, \varrho)$ assez petite. Dès lors, il suffit de poser $\Phi(\varepsilon \sin t) = F(t)$ avec un ε assez petit pour se ramener au cas précédent. On en déduit que Φ est analytique au voisinage de 0, d'ou le résultat annoncé.

La méthode se prète à des variantes. Signalons les principales.

1. Au lieu de prendre pour ensemble de définition de Φ l'intervalle ouvert $]-1,1[$, il est naturel de prendre un ensemble E du plan de la variable complexe $x+iy$: on dira que Φ opère dans $A(\Gamma)$ si, pour toute $f \in A(\Gamma)$ à valeurs dans E, $\Phi \circ f \in A(\Gamma)$. Si E est convexe et fermé (et si Γ est supposé, pour simplifier, compact), il faut et suffit, pour que Φ opère, que Φ soit analytique-réelle, en x et y, dans un voisinage de E [19]. A notre connaissance, il n'y a pas d'autre résultat dans cette direction.

2. Au lieu de considérer l'algèbre $A(\Gamma)$, on peut considérer l'algèbre $B(\Gamma)$ des transformées de Fourier de mesures. Evidemment, $B(\Gamma) = A(\Gamma)$ si Γ est compact. En dehors de ce cas, on établit que si Φ, définie sur $]-1,1[$, opère dans $B(\Gamma)$, c'est une fonction entière [37, 20]. Ce résultat est lié à l'asymétrie de l'algèbre $B(\Gamma)$. Un champ de recherches ouvert est la classification des idéaux de $B(\Gamma)$, contenant $A(\Gamma)$, suivant l'ensemble des fonctions qui y opèrent; quelques suggestions dans ce sens sont faites dans le livre de Rudin [69].

3. Dans le cas ou Γ est discret (disons, $\Gamma = Z$), on peut, au lieu de $A(\Gamma)$ considérer un idéal de $A(\Gamma)$, c'est à dire l'ensemble I_E des $f \in A(\Gamma)$ qui s'annulent sur une partie E de Γ. Pour des ensembles E assez lacunaires, il résulte d'un théorème de Zygmund [83] que les fonctions Φ lipschitziennes en 0 opèrent, et elles seules. Au contraire, pour des ensembles E contenant des progressions arithmétiques arbitrairement longues, les fonctions analytiques et elles seules opèrent [19]. Rien n'est connu entre ces extrèmes.

4. Dans tous les cas, on peut poser le problème en remplaçant $A(\Gamma)$ par une algèbre quotient $A(E) = A(\Gamma)/I_E$. Prenons pour simplifier $\Gamma = T$. Si E est un ensemble de Helson, toutes les fonctions continues opèrent. Par contre, si E contient des progressions arithmétiques arbitrairement longues

[36], ou même seulement si E contient des ν-mailles pour des ν arbitrairement grands, seules les fonctions analytiques opèrent [42]. Katznelson conjecture [43] qu'à part le cas des ensembles de Helson, les seules fonctions qui opèrent dans $A(E)$ sont les fonctions analytiques.

5. Enfin, au lieu de l'hypothèse que, pour toute $f \in A(T)$ à valeurs dans l'ensemble de définition de Φ, on a $\Phi \circ f \in A(T)$, Rudin a considéré l'hypothèse plus faible que, pour toute f, il existe un $p = p(f) < 2$ tel que la transformée de Fourier de $\Phi \circ f$ appartienne à $L^p(Z)$: la conclusion vaut encore [68].

Une autre direction, qui devait s'avérer très fructueuse, fut prise par P. Malliavin [53, 54]. Au lieu de supposer, comme tout à l'heure, que Φ opère dans $A(T)$, supposons seulement que Φ opère sur un élement f donné; autrement dit, que $\Phi \circ f \in A(T)$. Le théorème de Malliavin est que, pour certaines fonctions $f \in A(T)$, la simple hypothèse $\Phi \circ f \in A(T)$ entraîne que Φ est indéfiniment dérivable, et même que Φ appartient à une classe de fonctions indéfiniment dérivables d'un certain type [54].

L'idée de Malliavin est de considérer e^{iuf} non plus comme un élément de $A(T)$, mais comme une pseudomesure (élément du dual $A^*(T)$) et de chercher, non plus à minorer $\|e^{iuf}\|_A$, mais à majorer $\|e^{iuf}\|_{A^*}$. En fait, en prenant pour f une série de Fourier très lacunaire, on voit que $\|e^{iuf}\|_{A^*}$ est une fonction rapidement décroissante, soit $T(u)$. Si maintenant on applique la pseudomesure e^{iuf} à la fonction $\Phi \circ f \in A(T)$, on trouve

$$\left| \int_T \Phi(f(x)) e^{iuf(x)}\, dx \right| \leqslant \|\Phi \circ f\|_A \, \|e^{iuf}\|_{A^*} = \|\Phi \circ f\| T(u)$$

et aussi bien on a l'inégalité

$$\left| \int_T e^{iuf(x)}\, dx \right| \leqslant T(u).$$

Mais, en désignant par $d\mu(t)$ l'image de la mesure de Lebesgue par f, ces intégrales s'écrivent respectivement

$$\int \Phi(t) e^{iut}\, d\mu(t) \quad \text{et} \quad \int e^{iut}\, d\mu(t)\ ;$$

ce sont les transformées de Fourier de $\Phi(t)\, d\mu(t)$ et de $d\mu(t)$. Du fait que $T(u)$ est rapidement décroissante résulte que $d\mu(t) = \varrho(t) dt$, avec $\varrho \in C^\infty$, et que $\Phi\varrho \in C^\infty$. Quitte à vérifier que ϱ ne s'annule pas sur un certain intervalle, on voit que Φ y est indéfiniment dérivable.

Nous allons voir dans un instant le succès de cette idée dans un domaine *a priori* tout différent.

6. Les idéaux fermés de $A(\Gamma)$ et la synthèse harmonique dans $L^\infty(G)$

Il s'agit encore d'un problème ancien, qui remonte à Wiener et Beurling [4]. On sait que, si Γ est discret, tout idéal fermé de $A(\Gamma)$ est de la forme I_E (idéal fermé des $f \in A(\Gamma)$ qui s'annulent sur un fermé $E \subset \Gamma$). En est-il de même si Γ n'est pas discret ?

Par dualité, le problème s'énonce ainsi (« synthèse harmonique ») : est-il vrai que tout sous-espace faiblement fermé de $L^\infty(G)$ (considéré comme

dual de $L^1(G)$), invariant par translation, soit engendré par les caractères qu'il contient ?

On peut encore l'exprimer de la manière suivante : soit E un fermé $\subset\Gamma$, f une fonction $\in A(\Gamma)$ nulle sur E, T une pseudomesure à support dans E. S'ensuit-il que $(T,f)=0$? L'hypothèse était plausible du fait qu'il en est bien ainsi si T est une mesure.

Un contre-exemple très simple fut donné par L. Schwartz en 1947 [71], dans le cas $\Gamma=R^3$: E était une sphère, T la dérivée dans une direction de la mesure équirépartie sur la sphère (et toute l'idée du contre-exemple tient au fait que c'est bien une pseudomesure), et f une fonction quelconque, nulle sur E, telle que $(T,f)\neq 0$. Dès lors, l'intérêt se porta surtout sur le cas $\Gamma=R$ ou T.

Une série de conditions suffisantes fut donnée, liant E, f, T (f nulle sur E, T portée par E), pour assurer que $(T,f)=0$. Les plus intéressantes portent sur f seule (conditions pour que f « satisfasse la synthèse harmonique ») ou sur E seul (conditions pour E soit « un ensemble de synthèse harmonique »).

Comme exemple des premières, citons le résultat de Beurling-Pollard [59] : si $f\in A(T)$ est lipschitzienne d'ordre $\frac{1}{2}$ ($\omega_f(\delta)=0(\delta^{\frac{1}{2}})$, elle satisfait la synthèse harmonique.

Comme exemples des secondes : si la frontière de E est dénombrable, E est un ensemble de synthèse harmonique [10, 56]; de même (C. S. Herz) si E est le cercle dans R^2 [22] ou l'ensemble triadique de Cantor sur R[21], ou la frontière d'un polyèdre [23], ou un ensemble sans vraie pseudomesure (évident).

En 1959, Malliavin apporta la réponse longtemps attendue : l'hypothèse de la synthèse harmonique est fausse pour $\Gamma=\mathrm{R}$, et plus généralement pour tout groupe Γ non-discret [51, 52, 53]. La démonstration est en étroit rapport avec son travail sur le calcul symbolique. En effet, soit comme tout à l'heure f une fonction réelle $\in A(T)$, telle que les normes $\|e^{iuf}\|_{A^*}$ soient rapidement décroissantes ($u\to\infty$) et telles que de plus la densité ϱ de l'image de la mesure de Lebesgue par f ne s'annule pas sur un intervalle I contenant 0. Soit $[f]$ la sous-algèbre de $A(T)$ constituée par les fonctions de la forme $\Phi\circ f$. Des inégalités écrites tout à l'heure résulte aisément la continuité de l'application $\Phi\circ f\to\Phi$, de $[f]$ dans l'espace $C^1(I)$ des fonctions contínûment dérivables sur I. Toute forme linéaire θ sur $C^1(I)$ définit une forme linéaire sur $[f]$, que l'on peut étendre à $A(T)$: cette extension, que nous convenons de noter θ_f, est une pseudomesure telle que $(\theta_f,f)=(\theta_t,t)$ (valeur de la distribution θ sur la fonction $t\to t$); on peut d'ailleurs la choisir de telle sorte qu'elle soit portée par l'image réciproque du support de θ par f. Pour avoir le contre exemple cherché, il suffit maintenant de prendre pour θ la dérivée δ' de la mesure de Dirac, et $T=\delta'_f$: c'est bien une pseudomesure portée par l'ensemble des zéros de f, et $(T,f)=(\delta'_t,t)=-1$.

Le théorème de Malliavin donna immédiatement naissance à d'assez nombreux travaux [31, 47, 32, 66, 69]. Comme la construction faite par Malliavin de la fonction f est assez laborieuse, il a paru intéressant de la remplacer par un argument probabiliste : si l'on considère les fonctions aléatoires

$$f(x)=\sum_0^\infty r_n(X_{2n}\cos 2\pi nt+X_{2n+1}\sin 2\pi nt),$$

ou r_n désigne une suite positive sommable astreinte à un certaine condition (par exemple, $r_{4^j}>1/j^2$; $j=1,2,\ldots$) et ou les X_n sont des variables aléatoires

gaussiennes normales, indépendantes les unes des autres, alors il est presque sûr que f ne satisfait pas la synthèse harmonique. Citons quelques résultats qu'on peut obtenir ainsi :

— il existe des $f \in A(T)$ satisfaisant une condition de Lipchitz d'ordre $\alpha > 0$, et qui ne satisfont pas la synthèse harmonique. Signalons que pour des raisons techniques, on est resté bloqué à $\alpha < \frac{1}{4}$, alors que la borne supérieure des α est sans doute $\frac{1}{2}$ [31], [32].

— il existe des $f \in A(T)$ tels que les idéaux fermés engendrés par les puissances f^n $(n = 1, 2, \ldots)$ soient tous différents [66]; il existe des $f \in A(T)$ tels que les idéaux engendrés par f et $\bar{f}$ (conjugué de f) soient différents [68].

— alors que toute pseudomesure sur T dont la transformée de Fourier appartient à $L^2(Z)$ satisfait la synthèse harmonique (puisque c'est une mesure), il existe des pseudomesures que ne satisfont pas la synthèse harmonique, et qui appartiennent à $L^p(Z)$ pour tout $p > 2$ [32].

Ce qui reste de plus mystérieux est la structure des ensembles de synthèse harmonique. Et c'est aussi le plus intéressant pour la classification des idéaux des algèbres $A(\Gamma)$. En effet, si E est un ensemble de synthèse harmonique, il existe un seul idéal fermé, à savoir I_E, tel que E soit l'ensemble des zéros communs aux fonctions qui le composent. Tandis que si E n'est pas un ensemble de synthèse harmonique, un travail déjà ancien de Helson [15] établit qu'il existe une infinité de tels idéaux.

L'argument probabiliste invoqué tout à l'heure montre que, dans un certain sens, il y a beaucoup d'ensembles qui ne sont pas de synthèse harmonique. Et il est frappant que, jusqu'à présent, on n'ait pas réussi à en désigner un seul parmi les ensembles qui s'imposent à l'attention (ensembles du type de Cantor, etc...).

Et particulier, est-il vrai que tout ensemble de Helson soit ensemble de synthèse harmonique ? La question revient à celle, que nous avons déjà posée, de savoir s'il existe d'autres ensembles de Helson que les ensembles sans vraie pseudomesure [38].

La difficulté d'une classification des ensembles fermés suivant qu'ils sont, ou non, ensembles de synthèse harmonique, apparait bien par la remarque suivante : étant donné un fermé quelconque, il suffit de lui ajouter un dénombrable convenable pour le transformer en ensemble de synthèse harmonique (cf. p. ex. [39]).

Pour éviter cette difficulté, Malliavin a introduit tout récemment la notion d'ensemble de résolution spectrale [55]. Il s'agit des ensembles fermés dont toute partie fermée est ensemble de synthèse harmonique. Il n'est pas exclu que cette classe d'ensembles coincide avec celle des fermés qui sont « ensembles d'unicité » pour le développement trigonométrique (c'est à dire qui ne portent aucune distribution $\neq 0$ dont les coefficients de Fourier soient nuls à l'infini). Malliavin démontre l'inclusion dans un sens (tout ensemble de résolution est ensemble d'unicité).(1)

7. Les sous-algèbres fermées de $A(\Gamma)$

Lorsque Γ n'est pas discret, il existe déjà tant d'idéaux fermés dans $A(\Gamma)$ que la classification des sous-algèbres fermées apparait comme hors de portée.

(1) (Note sur épreuves.) Un travail à paraître, de Kahane et Katznelson, dément la conjecture : il existe des ensembles d'unicité (par exemple l'ensemble triadique de Cantor) qui ne sont pas de résolution.

Bornons nous donc d'abord au cas ou Γ est discret et, pour préciser les idées, prenons $\Gamma = Z$. Pour obtenir une sous-algèbre fermée de $A(Z)$, on peut procéder ainsi : on considère une partition $\mathcal{P}$ de Z contenant au plus une classe infinie, et l'ensemble $\bar{A}(\mathcal{P})$ de toutes les $f \in A(Z)$ qui sont constantes sur chaque classe de $\mathcal{P}$. $\bar{A}(Z)$ est une sous-algèbre fermée. Une question naturelle, posée par Rudin dans son livre, est la suivante : obtient-on ainsi toutes les sous-algèbres fermées de $A(Z)$? On peut encore l'exprimer ainsi : est-il vrai que, pour chaque partition $\mathcal{P}$, l'idéal de $\bar{A}(\mathcal{P})$ engendré par les f à support fini, soit $\underline{A}(\mathcal{P})$, coincide avec $\bar{A}(\mathcal{P})$?

La réponse est négative, comme on l'a montré récemment [35]. D'autre part, il est facile de voir que, si les diamètres des classes finies de $\mathcal{P}$ sont bornés, on a $\underline{A}(\mathcal{P}) = \bar{A}(\mathcal{P})$. Une classification des partitions, suivant que cette égalité a lieu ou non, serait souhaitable, mais elle apparait difficile.

D'autres recherches intéressantes sur les sous-algèbres fermées des algèbres $A(\Gamma)$ ont été menées récemment, suivant trois directions principales.

1. Malliavin, comme corollaire de ses résultats sur le calcul symbolique, a montré que, si Γ est compact, il existe des $f \in A(\Gamma)$ telle que la sous-algèbre fermée engendrée par f, soit (f), soit strictement contenue dans la sous-algèbre $[f]$ formée par les éléments de $A(\Gamma)$ de la forme $\Phi \circ f$. Suivant la méthode de Marcinkiewicz, il a aussi donné des conditions suffisantes, portant sur Ψ, pour que $\Psi \circ f \in (f)$ [54].

2. J. Wermer et A. B. Simon ont étudié et élucidé le problème suivant [78, 79, 75, 76] : quels sont les groupes G, et les parties S de G, tels que l'ensemble des $f \in A(\Gamma)$ qui s'écrivent

$$f(x) = \int_S \hat{f}(u)\,(x, u)\,dm(u)$$

forme une sous-algèbre maximale de $A(\Gamma)$? Wermer montre qu'il en est ainsi si $G = R$, et $S = (-\infty, 0]$, ou si G est un sous-groupe discret des réels, et $S = G \cap (-\infty, 0]$. Simon montre que ce sont substantiellement les seuls cas possibles.

3. Katznelson et Rudin ont cherché s'il existe des sous-algèbres fermées de $A(\Gamma)$, autoadjointes (c'est à dire contenant, en même temps que f, la fonction conjuguée $\bar{f}$), séparant les points de Γ_∞ (rappelons que $f(\infty) = 0$), et distinctes de $A(\Gamma)$. Le problème est suggéré par le fait que, si l'on considère, au lieu de $A(\Gamma)$, l'algèbre des fonctions continues sur Γ (et éventuellement nulles à l'infini), la réponse est négative en vertu du théorème de Stone-Weierstrass : on dira donc que $A(\Gamma)$ est une « algèbre de Stone-Weierstrass » dans le cas ou la réponse est négative. Le résultat est la suivant : $A(\Gamma)$ est une algèbre de Stone-Weierstrass si et seulement Γ est totalement discontinu [44, 69]. En particulier, $A(T)$ n'est pas une algèbre de Stone-Weierstrass.

8. Conclusion

Quoique que la revue que nous avons faite des résultats récents — et moins récents — concernant les classes $A(\Gamma)$ soit bien incomplète, elle donnera sans doute l'impression d'être une collection variée de résultats et de problèmes, plus que l'exposé d'une théorie. Cela tient peut-être à ce que

ces classes $A(\Gamma)$ forment un matériau de construction qui se prête à plusieurs types d'architecture, et que plusieurs théories peuvent trouver à s'y appliquer. Elles ressortissent aussi bien à ce que S. Bernstein appelait la théorie constructive des fonctions, qu'à celle des algèbres de Banach, ou à l'analyse de Fourier classique. Au cours de ces dernières années, nous pensons avoir montré qu'elles ont suggéré beaucoup de bon travail, et aussi que beaucoup de problèmes restent non résolus. Au terme de ce trop long exposé, nous émettons le voeu que de nouvelles énergies s'y attaquent.

BIBLIOGRAPHIE

[1]. Banach, S., Über einige Eigenschaften der lakunären trigonometrischen Reihen. *Studia Math.*, 2 (1930), 207–220.

[2]. Bernstein, S., Sur la convergence absolue des séries trigonométriques. *C. R. Acad. Sci. Paris*, 158 (1914), 1661–1663.

[3]. —— Sur la convergence absolue des séries trigonométriques. *C. R. Acad. Sci. Paris*, 199 (1934), 397–400.

[4]. Beurling, A., Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle. *Neuvième Congrès Math. Scand. Helsinki*, 1938, 345–366. Helsingfors 1939.

[5]. —— On the spectral synthesis of bounded functions. *Acta Math.*, 81 (1949), 225–238.

[6]. —— Séminaire tenu en 1959 à l'Institute for Advanced Study, Princeton. (Non-publié.)

[7]. Beurling, A. & Helson, H., Fourier-Stieltjes transforms with bounded powers. *Math. Scand.*, 1 (1953), 120–126.

[8]. Boas, R. P., Beurling's test for absolute convergence of Fourier series. *Bull. Amer. Math. Soc.*, 66 (1960), 24–26.

[9]. Carleson, L., Sets of uniqueness for functions regular in the unit circle. *Acta Math.*, 87 (1952), 325–345.

[10]. Щцлов, Г. Е., *О регулярных нормироьаппых кольцах Труци мат. иист. Стеклов*, 21 (1947).

[11]. Cohen, P. J., On a conjecture of Littlewood and idempotent measures. *Amer. J. Math.*, 82 (1960), 191–212.

[12]. —— On homomorphisms of group algebras. *Amer. J. Math.*, 82 (1960), 213–226.

[13]. Davenport, H., On a theorem of P. J. Cohen. *Mathematica*, 7 (1960), 93–97.

[14]. Helson, H., Spectral synthesis of bounded functions. *Ark. Mat.*, 1 (1951), 497–502.

[15]. —— On the ideal structure of group algebras. *Ark. Mat.*, 2 (1952) 83–86.

[16]. —— Isomorphisms of Abelian group algebras. *Ark. Mat.*, 2 (1953), 475–487.

[17]. —— Note on harmonic functions. *Proc. Amer. Math. Soc.*, 4 (1953), 686–691.

[18]. —— Fourier transforms on perfect sets. *Studia Math.*, 14 (1954), 209–213.

[19]. Helson, H. & Kahane, J.-P., Sur les fonctions opérant dans les algèbres de transformées de Fourier de suites ou de fonctions sommables. *C.R. Acad. Sci. Paris*, 247 (1958), 626–628.

[20]. Helson, H., Kahane, J.-P., Katznelson, Y. & Rudin, W., The functions which operate on Fourier transforms. *Acta Math.*, 102 (1959), 135–157.

[21]. HERZ, C. S., Spectral synthesis for the Cantor set. *Proc. Nat. Acad. Sci. U.S.* 42 (1956), 42–43.

[22]. —— Spectral synthesis for the circle. *Ann. Math.* 68 (1958), 709–712.

[23]. —— The spectral theory of bounded functions. *Trans. Amer. Math. Soc.* 94 (1960), 181–232.

[24]. HEWITT, E. & ZUCKERMAN, H. S., Some theorems on lacunary Fourier series, with extensions to compact groups. *Trans. Amer. Math. Soc.* 93 (1959), 1–19.

[25]. ИВАТЕВ-МУСАТОВ, О. С., О коэффициентах тригонометрических нуль-рядов. *Изв. А. Н. С.С.С.Р.*, 21 (1957), 559–578.

[26]. KAHANE, J.-P., Sur certaines classes de séries de Fourier absolument convergentes. *J. Math. Pures Appl.* 35 (1956), 249–259.

[27]. —— Généralisation d'un théorème de S. Bernstein. *Bull. Soc. Math. France*, 85 (1957), 221–229.

[28]. —— Sur un problème de Littlewood. *Nederl. Akad. Wetensch., Indag. Math.*, 19 (1957), 268–271.

[29]. —— Sur les fonctions moyenne-périodiques bornées. *Annales Institut Fourier*, 7 (1958). 293–314.

[30] —— Sur un théorème de Wiener-Lévy. *C. R. Acad. Sci. Paris*, 246 (1958), 1949–1951.

[31]. —— Sur un théorème de Paul Malliavin. *C. R. Acad. Sci. Paris*, 248 (1959), 2943–2944.

[32]. —— Sur la synthèse harmonique dans l^∞. *Anaïs de Academia Brasileira de Ciencias*, 32 (1960).

[33]. —— Algebras de convolucion de sucesiones, funciones y medidas sumables (notes d'un cours donné en 1959). *Cursos y seminarios de matematica, Buenos-Aires*, 6 (1961).

[34]. —— Fonctions pseudopériodiques et séries de Fourier lacunaires. *Ann. Sci. E.N.S.*, 79 (1962), 93–150.

[35]. —— Sur les sous-algèbres fermées de $L^1(T)$. *C. R. Acad. Sci. Paris*, 254 (1962), 4415–4416.

[37]. KAHANE, J.-P. & KATZNELSON, Y., Sur la réciproque du théorème de Wiener-Lévy. *C. R. Acad. Sci. Paris*, 248 (1959) 1279–1281.

[37]. KAHANE, J.-P. & RUDIN, W., Caractérisation de fonctions qui opèrent sur les coefficients de Fourier-Stieltjes. *C. R. Acad. Sci. Paris*, 247 (1958), 773–775.

[38]. KAHANE, J.-P. & SALEM, R., Sur les ensembles linéaires ne portant pas de pseudomesures. *C. R. Acad. Sci. Paris*, 243 (1956), 1185–1187. Sur les ensembles de Carleson et de Helson. *C. R. Acad. Sci. Paris*, 243 (1956), 1706–1708.

[39]. —— Ensembles parfaits et séries trigonométriques. *Actualités Sci. Ind.*, Hermann (Paris), 1963.

[40]. KATZNELSON, Y., Sur les fonctions opérant sur l'algèbre des séries de Fourier absolument convergentes. *C. R. Acad. Sci. Paris*, 247 (1958), 404–406.

[41]. —— Sur le calcul symbolique dans quelques algèbres de Banach. *Ann. Sci. E.N.S.*, 76 (1959), 83–124.

[42]. —— Sur un théorème de Kahane et Salem concernant les ensembles symétriques. *C. R. Acad. Sci. Paris*, 254 (1962), 2514–2515.

[43]. —— Calcul symbolique dans les algèbres homogènes. *C. R. Acad. Sci. Paris*, 254 (1962), 2700–2702.

[44]. KATZNELSON, Y. & RUDIN, W., The Stone-Weierstrass property in Banach algebras. *Pacific J. Math.*, 11 (1961), 253–265.

[45]. KINUKAWA, M., Contractions of Fourier coefficients and Fourier integrals. *J. Analyse Math.*, 8 (1960–61), 377–406.

[46]. LANDAU, H., (Travail non publié).

[47]. KOOSIS, P., Sur un théorème remarquable de Paul Malliavin. *C. R. Acad. Sci. Paris*, 249 (1959).

[48]. ЛЕЙЬЕНЗОН, З. Л., О кольйе функций с абсолютно сходящимися рядами Фурье. *Усп. мат. наук*, 9 (1954), 157–162.

[49]. LEVY, P., Sur la convergence absolue des séries de Fourier. *Compositio Math.*, 1 (1934), 1–14.

[50]. LITTLEWOOD, J. E., On mean values of power series. *Proc. London Math. Soc.*, 25 (1926), 328–37.

[51]. MALLIAVIN, P., Sur l'impossibilité de la synthèse spectrale dans une algèbre de fonctions presque périodiques. *C. R. Acad. Sci. Paris*, 248 (1959), 1756–1759.

[52]. —— Sur l'impossibilité de la synthèse spectrale sur la droite. *C. R. Acad. Sci. Paris*, 248 (1959), 2155–2157.

[53]. —— Impossibilité de la synthèse spectrale sur les groupes abéliens non compacts. *Publ. Math. (I.H.E.S.), Paris*, 1959, 61–68.

[54]. —— Calcul symbolique et sous-algèbres de $L_1(G)$. *Bull. Soc. Math. France*, 87 (1959), 181–190.

[55]. —— Ensembles de résolution spectrale. *Proc. Internat. Congress Math. Stockholm*, 1962, 368–378. Uppsala 1963.

[56]. MANDELBROJT, S. & AGMON, S., Une généralisation du théorème taubérien de Wiener. *Acta Litt. Sci. Szeged*, 12 (1950), 167–176.

[57]. MARCINKIEWICZ, J., Sur la convergence absolue des series de Fourier. *Mathematica, Cluj*, 16 (1940), 66–73.

[58]. PALEY, R. E. & ZYGMUND, A., On some series of functions. (I). *Proc. Cambridge Phil. Soc.*, 26 (1930), 337–57.

[59]. POLLARD, H., The harmonic analysis of bounded functions. *Duke Math. J.*, 20 (1953), 499–512.

[60]. RUDIN, W., The automorphisms and the endomorphisms of the group algebra of the unit circle. *Acta Math.*, 95 (1956), 39–56.

[61]. —— Transformations des coefficients de Fourier. *C. R. Acad. Sci. Paris*, 243 (1956), 638–640.

[62]. —— On isomorphisms of group algebras. *Bull. Amer. Math. Soc.*, 64 (1958), 167–169.

[63]. —— Idempotent measures on Abelian groups. *Pacific J. Math.*, 9 (1959), 195–209.

[64]. —— Measure algebras on Abelian groups. *Bull. Amer. Math. Soc.*, 65 (1959), 227–247.

[65]. —— Trigonometric series with gaps. *J. Math. Mech.*, 9 (1960), 203–228.

[66]. —— Closed ideals in group algebras. *Bull. Amer. Math. Soc.*, 66 (1960), 81–83.

[67]. —— Fourier-Stieltjes transforms of measures on independent sets. *Bull. Amer. Math. Soc.*, 66 (1960), 199–202.

[68]. —— A strong converse of the Wiener-Lévy theorem. *MRC Technical Summary Report*. No. 261. Madison, Wisc., 1961.

[69]. —— *Fourier transforms on groups*. Interscience Publishers, New York, 1962.

[70]. SALEM, R., On singular monotonic functions whose spectrum has a given dimension. *Ark. Mat.*, 1 (1951), 353–365.

[71]. SCHWARTZ, L., Sur une proprieté de synthèse spectrale dans les groupes non compacts. *C. R. Acad. Sci. Paris*, 227 (1948), 424–426.

[72]. SEGAL, I. E., The class of functions which are absolutely convergent Fourier transforms. *Acta Litt. Sci. Szeged*, 12 (1950), 157–161.

[73]. SIDON, S., Verallgemeinerung eines Satzes über die absolute Konvergenz von Fourierreihen mit Lücken. *Math. Ann.*, 97 (1927), 675–676.

[74]. —— Einige Sätze und Fragestellungen über Fourier-Koeffizienten. *Math. Z.*, 34 (1932), 477–480.

[75]. SIMON, A. B., Vanishing algebras. *Trans. Amer. Math. Soc.*, 92 (1959), 154–167.

[76]. —— On the maximality of vanishing algebras. *Amer. J. Math.*, 81 (1959), 613–616.

[77]. СТЕЧКИН, С. Д., Об абсолютной сходимости рядов Фурье (I) *изв. А.н. С.С.С.Р*, сер. мат. 17 (1953), 87–98. (II) *Ibid.*, 19 (1955), 221–246. (III) *Ibid.*, 20 (1956), 385–412.

[78]. WERMER, J., Maximal subalgebras of group algebras. *Proc. Amer. Math. Soc.*, 6 (1955), 692–694.

[79]. —— On a subalgebra of $L(-\infty, \infty)$. *Amer. J. Math.*, 82 (1960), 103–105.

[80]. WIENER, N., Tauberian theorems. *Ann. Math.*, 33 (1932), 1–100.

[81]. WIENER, N., & WINTNER, A., On singular distributions, *J. Math. Phys.* 17 (1938), 233–246.

[82]. WIK, I., On linear dependence in closed sets. *Ark. Mat.*, 4 (1960), 209–218.

[83]. ZYGMUND, A., Sur les séries trigonométriques lacunaires. *J. London Math. Soc.*, 5 (1930), 138–145.

[84]. —— *Trigonometric Series*, 2nd ed., vols. I, II. Cambridge Univ. Press, 1959.

TOPOLOGICAL MANIFOLDS AND SMOOTH(1) MANIFOLDS

By J. MILNOR

Suppose that one is given a topological manifold M (i.e. a Hausdorff space with a countable basis where each point has a neighborhood homeomorphic to some euclidean space). Then one can ask the following two questions:

Problem 1. Can M be given the structure of a smooth manifold? In more intuitive terms: can M be imbedded in a high dimensional euclidean space so as to have a continuously turning tangent plane?

Problem 2. If such a smoothness structure exists, is it essentially unique? More precisely, given two such structures on M, does there exist a homeomorphism of M onto itself which carries one structure to the other?(2)

The first problem was answered negatively when M. Kervaire gave an example of a compact triangulable 10-dimensional manifold which is not smoothable. (Thus if Kervaire's manifold is imbedded in some euclidean space, its image must have "angles" or "corners" or worse singularities.) Other such examples, in other dimensions, have been given by Smale, Tamura, Wall and by Eells and Kuiper (references [5], [6], [14], [18], [21]).

The second problem was answered negatively when the author showed that the 7-dimensional sphere possesses several essentially distinct smoothness structures (see [8], [9], [13], [17]).

Thus the two problems are non-trivial. They lead naturally to the following.

Problem 3. Given a topological manifold M, can one make a classification of all possible smoothness structures on M?

The answer must surely depend on a detailed knowledge of the topology of M.

Quite a bit of progress on these questions has been made during the last few years. Suppose for example that M is the topological sphere S^n. Define two smoothness structures on S^n to be *equivalent* if there exists an orientation preserving homeomorphism of S^n to itself which carries one smoothness structure to the other. For $n \neq 4$ it is known that the set of such equivalence classes can be made into an abelian group, which is denoted by Γ_n. The structure of this group for many small values of n has been determined by Kervaire and Milnor, making use of work by Smale (references [7], [10], [14], [15]). (For the cases $n < 4$ see [11], 22].) For example one has:

$$\Gamma_1 = \Gamma_2 = \Gamma_3 = \Gamma_5 = \Gamma_6 = 0, \quad \Gamma_7 = Z_{28}, \quad \Gamma_8 = Z_2.$$

The groups Γ_n, $n \neq 4$, are all finite.(3)

(1) The word *smooth* will be used as a synonym for "differentiable of class C^∞".

(2) This is equivalent to the question as to whether the two resulting smooth manifolds are diffeomorphic to each other.

(3) Here Z_k denotes the cyclic group of order k. For $n = 4$ the group Γ_4 must be defined somewhat differently. Nothing is known about the structure of Γ_4.

Now let M be an arbitrary triangulated manifold. J. Munkres, in reference [12], has defined a sequence of obstructions, whose vanishing implies that M can be given a smoothness structure (see also Thom [19], [20]). These obstructions are homology classes of M with coefficients in the groups Γ_i. Similarly, if one is given two different smooth manifolds with the same underlying complex, Munkres [11] has defined a sequence of obstruction classes whose vanishing implies that the two manifolds are diffeomorphic. Again the groups Γ_i occur as coefficient groups.

One interesting application of Munkres' results has been made by J. Stallings. In reference [16], Stallings shows that the euclidean space $R^n, n \neq 4$, has an essentially unique smoothness structure.

In the remainder of this lecture, I would like to introduce a quite different tool, which I hope will be used in the future to attack these problems; namely the theory of microbundles.

A "microbundle" is an object something like a fibre bundle having the euclidean space R^n as fibre. However the fibre in a microbundle is not an honest topological space, but is only a "germ" of a topological space. This can be made precise as follows.

Definition. An R^n-*microbundle* over B is a commutative diagram

$$\begin{tikzcd} & E \arrow[dr, "j"] & \\ B \arrow[ur, "i"] \arrow[rr, "\text{identity}"] & & B \end{tikzcd}$$

where B, E are topological spaces, and i, j are continuous maps; such that the following "local triviality" condition is satisfied:

Requirement. For each $b \in B$ there should exist neighborhoods U of b and V of $i(b)$, with

$$i(U) \subset V, \; j(V) = U$$

so that V is homeomorphic to $U \times R^n$ under a homemorphism which makes the following diagram commutative

$$\begin{tikzcd} & V \arrow[dr, "j"] \arrow[dd] & \\ U \arrow[ur, "i"] \arrow[rr] \arrow[dr, "\times 0"'] & & U \\ & U \times R^n \arrow[ur, "\mathrm{proj}_1"'] & \end{tikzcd}$$

Here R^n denotes the n-dimensional euclidean space, $\times 0$ denotes the mapping $u \to (u, 0)$, and proj_1 denotes the projection to the first factor: $\mathrm{proj}_1\,(u, x) = u$.

Such a microbundle will be denoted by a single German letter, such as $\mathfrak{x}$. The spaces E and B will be called the *total space* and the *base space* respectively. The maps i, j will be called the *injection* and the *projection* maps of $\mathfrak{x}$.

Note that this condition of local triviality depends only on that portion of E which lies in a aribitrarily small neighborhood of $i(B)$. If E_0 is any neighborhood of $i(B)$ in E then we will take the point of view that the new microbundle

$$\begin{tikzcd} & E_0 \arrow[dr, "j|E_0"] & \\ B \arrow[ur, "i"] \arrow[rr, "\text{identity}"] & & B \end{tikzcd}$$

can be identified with the original one. More precisely, and more generally:

Definition. A second microbundle $\mathfrak{x}'$ over B with diagram

$$\begin{tikzcd} & E' \arrow[dr, "j'"] & \\ B \arrow[ur, "i'"] \arrow[rr, "\text{identity}"] & & B \end{tikzcd}$$

is *isomorphic* to $\mathfrak{x}$ if there exist neighborhoods E_0 of $i(B)$ in E and E'_0 of $i'(B)$ in E', and a homeorphism from E_0 to E'_0 which makes the following diagram commutative.

$$\begin{tikzcd} & E \arrow[dr, "j"] & \\ & E_0 \arrow[d] & \\ B \arrow[ur, "i"] \arrow[r] \arrow[dr, "i'"'] & B & \\ & E'_0 \arrow[u] & \\ & E' \arrow[u, phantom, "\cap"] \arrow[uur, "j'"'] & \end{tikzcd}$$

Here are some examples of microbundles.

Example 1. For any topological space B and any integer $n \geqslant 0$ one has the *trivial* microbundle $\mathfrak{e}^n$ with diagram

$$\begin{tikzcd} & B \times R^n \arrow[dr, "\text{proj}_1"] & \\ B \arrow[ur, "\times 0"] \arrow[rr, "\text{identity}"] & & B \end{tikzcd}$$

More generally any microbundle isomorphic to $\mathfrak{e}^n$ is called a trivial microbundle.

Example 2. Let ξ be a vector bundle over B with total space $E(\xi)$ and projection map $p: E(\xi) \to B$. There is a standard cross-section

$$z: B \to E(\xi)$$

which assigns to each $b \in B$ the zero vector in the vector space $p^{-1}(b)$. The *underlying microbundle* $|\xi|$ of ξ is defined to be the microbundle

$$\begin{tikzcd} & E(\xi) \arrow[dr, "p"] & \\ B \arrow[ur, "z"] \arrow[rr, "\text{identity}"] & & B \end{tikzcd}$$

Example 3. Let M be a topological manifold. Then the *tangent microbundle* $\mathfrak{t}$ of M is defined to be the microbundle

$$\begin{tikzcd} & M \times M \arrow[dr, "\mathrm{proj}_1"] & \\ M \arrow[ur, "\Delta"] \arrow[rr, "\mathrm{identity}"] & & M \end{tikzcd}$$

where Δ denotes the diagonal map. Thus the "fibre" over a point $x_0 \in M$ is the set of all pairs (x_0, y) where y ranges over an arbitrary neighborhood of x_0 in M. The local triviality condition can be verified as follows. Given $x_0 \in M$ let U be a neighborhood homeomorphic to R^n under a homeomorphism h, and let $V = U \times U$. Then V is homeomorphic to $U \times R^n$ under the homeomorphism,

$$f(u_1, u_2) = (u_1, h(u_2) - h(u_1)),$$

which makes the following diagram commutative.

$$\begin{tikzcd} & V \arrow[dr, "\mathrm{proj}_1"] \arrow[dd, "f"] & \\ U \arrow[ur, "\Delta"] \arrow[rr, "\times 0"'] & & U \\ & U \times R^n \arrow[ur, "\mathrm{proj}_1"'] & \end{tikzcd}$$

Now suppose that M can be made into a smooth manifold. Then, using the smoothness structure, one can also define the *tangent vector bundle* τ of M. The following result is fundamental.

THEOREM 1. *In this situation the underlying microbundle* $|\tau|$ *is isomorphic to the tangent microbundle* $\mathfrak{t}$ *of* M.

The proof can be outlined as follows. Choose a Riemannian metric on M. Then for any tangent vector $v \in E(\tau)$ which is not too long, there exists a geodesic segment

$$\gamma_v : [0, 1] \to M,$$

whose velocity vector at 0 is the given vector v. Now the correspondence

$$v \to (\gamma_v(0), \gamma_v(1))$$

defines the required homeomorphism between a neighborhood of $z(M)$ in $E(\tau)$ and a neighborhood of the diagonal in $M \times M$.

COROLLARY. *If* M *can be smoothed then the tangent microbundle* $\mathfrak{t}$ *is isomorphic to* $|\xi|$ *for some vector bundle* ξ *over* M.

A fundamental conjecture would be the converse proposition:

Problem. If $\mathfrak{t}$ is isomorphic to $|\xi|$ for some ξ, does it follow that M can be given a smoothness structure?

The following partial result can be proved.

THEOREM 2. *If the tangent microbundle of* M *is isomorphic to* $|\xi|$ *for some vector bundle* ξ, *then the Cartesian product* $M \times R^{4n+1}$ *can be given a smoothness structure.*

I will not try to describe the proof, which is based on a method due to M. Curtis and R. Lashof [4].

Many standard constructions for vector bundles carry over immediately to microbundles. For example given a microbundle $\mathfrak{x}$ over B, and given a map $f: B' \to B$, one can construct the *induced microbundle* $f^*\mathfrak{x}$ over B'.

THEOREM 3. (Homotopy theorem.) *If B' is paracompact, and if $g: B' \to B$ is homotopic to f, then the induced bundle $g^*\mathfrak{x}$ is isomorphic to $f^*\mathfrak{x}$.*

The proof is similar to the usual proof for vector bundles.

Given two microbundles $\mathfrak{x}$ and $\mathfrak{y}$ over the same base space B, one can construct the *Whitney sum* $\mathfrak{x} \oplus \mathfrak{y}$, a new vector bundle over B. By definition, $\mathfrak{x} \oplus \mathfrak{y}$, is equal to $\Delta^*(\mathfrak{x} \times \mathfrak{y})$, where $\mathfrak{x} \times \mathfrak{y}$ denotes the Cartesian product microbundle

$$\begin{tikzcd} & E(\mathfrak{x}) \times E(\mathfrak{y}) \arrow[dr, "j \times j'"] & \\ B \times B \arrow[ur, "i \times i'"] \arrow[rr, "\text{identity}"] & & B \times B \end{tikzcd}$$

and where $\Delta: B \to B \times B$ denotes the diagonal map.

Definition. Two microbundles $\mathfrak{x}$ and $\mathfrak{x}'$ over B belong to the same s-class if there exist integers m, m' so that the Whitney sum $\mathfrak{x} \oplus \mathfrak{e}^m$ is isomorphic to $\mathfrak{x}' \oplus \mathfrak{e}^{m'}$. (Here $\mathfrak{e}^m$ denotes the trivial R^m-microbundle over B.)

THEOREM 4. *Let B be a finite dimensional complex. Then the s-classes of microbundles over B form an abelian group with respect to the Whitney sum operation.*

The proof is more difficult than the corresponding proof for vector bundles. The key step, showing that for each $\mathfrak{x}$ there exists a $\mathfrak{y}$ with $\mathfrak{x} \oplus \mathfrak{y}$ trivial, is proved by induction on the dimension of B.

This group of s-classes of microbundles will be denoted by $k_{\text{Top}}B$. The analogous group whose elements are s-classes of vector bundles over B will be denoted by $k_{\text{Orthog}}B$. Note that the correspondence $\xi \to |\xi|$ gives rise to a natural homomorphism

$$k_{\text{Orthog}}B \to k_{\text{Top}}B.$$

Note also that the groups $k_{\text{Top}}\ B$ (or $k_{\text{Orthog}}B$) behave somewhat like cohomology groups. Thus any map $f: B' \to B$ induces a homomorphism

$$f^*: k_{\text{Top}}B \to k_{\text{Top}}B'.$$

If f is a homotopy equivalence, then f^* is an isomorphism.

The groups $k_{\text{Orthog}}B$ are well known through the work of Atiyah, Hirzebruch, Adams and others (see [1], [2], [3]). Unfortunately very little is known about $k_{\text{Top}}B$. For example it is not known whether the groups $k_{\text{Top}}S^n$ are finite, countably infinite, or uncountably infinite. Even the group $k_{\text{Top}}S^1$ seems forbiddingly difficult to compute.

The following qualitative result can be obtained.

THEOREM 5. *There exists a finite complex X_1 for which the canonical homomorphism*

$$k_{\text{Orthog}}X_1 \to k_{\text{Top}}X_1$$

has a non-trivial kernel. Furthermore there exists a finite complex X_2 so that the canonical homomorphism

$$k_{\mathrm{Orthog}} X_2 \to k_{\mathrm{Top}} X_2$$

is not onto.

Thus the theory of microbundles is essentially distinct from the theory of vector bundles. The proof of Theorem 5 is quite difficult. It is based on joint research with M. Kervaire [7].

(Actually the proof of Theorem 5 gives a specific example of such a complex X_1: namely a 7-sphere with an 8-cell attached by a map of degree 7. For X_2 the proof shows only that one of two possibilities will work. If $k_{\mathrm{Top}} S^8$ is infinite then the 8-sphere itself will serve as a complex X_2. If $k_{\mathrm{Top}} S^8$ is finite, then the 8-sphere with a 9-cell attached by a map of degree 3 will serve.)

Each half of Theorem 5 has an interesting consequence.

COROLLARY 1. *The tangent vector bundle of a certain smooth manifold M_1 is not a topological invariant.*

Proof. Choose an open set U_1 in some euclidean space R^n which has the same homotopy type as X_1. Then there exists a vector bundle ξ over U_1 whose s-class is non-trivial, and belongs to the kernel of the homomorphism $k_{\mathrm{Orthog}} U_1 \to k_{\mathrm{Top}} U_1$. Thus the underlying microbundle $|\xi|$ is s-trivial. Without loss of generality we may assume that $|\xi|$ itself is trivial.

Let ε^p denote the trivial vector bundle, with total space $U_1 \times R^p$, where p is the fibre dimension of ξ. Since $|\varepsilon^p|$ is isomorphic to $|\xi|$ it follows that some neighborhood M_1 of $U_1 \times 0$ in $U_1 \times R^p$ is homeomorphic to some neighborhood M_1' of the zero cross-section in $E(\xi)$.

But each of the bundles ε^p and ξ can be given the structure of a smooth vector bundle. Hence the open sets $M_1 \subset E(\varepsilon^p)$ and $M_1' \subset E(\xi)$ can be considered as smooth manifolds. Clearly the manifold M_1 is parallelizable. However the tangent vector bundle of M_1', restricted to U_1, is isomorphic to

$$(\text{tangent bundle of } U_1) \oplus \xi \cong \varepsilon^n \oplus \xi.$$

Thus M_1' is not parallelizable. This completes the proof of Corollary 1.

COROLLARY 2. *There exists a topological manifold M_2 such that no Cartesian product $M_2 \times M'$ can be given a smoothness structure.*

Sketch of proof. Let U_2 be an open subset of some euclidean space R^n having the homotopy type of X_2. Then there exists a microbundle $\mathfrak{x}$ over U_2 whose s-class does not belong to the image of the homomorphism

$$k_{\mathrm{Orthog}} U_2 \to k_{\mathrm{Top}} U_2.$$

Let M_2 be the total space of this microbundle. We may assume that M_2 is a manifold.

It can be shown that the tangent microbundle of M_2, restricted to U_2, is isomorphic to the Whitney sum

$$(\text{tangent microbundle of } U_2) \oplus \mathfrak{x} \cong \mathfrak{e}^n \oplus \mathfrak{x}.$$

Since this is not isomorphic to $|\xi|$ for any vector bundle ξ over U_2, it follows from the Corollary to Theorem 1 that M_2 is not smoothable.

Given any positive integer p, a similar argument shows that the product $M_2 \times R^p$ is not smoothable. But this implies that no Cartesian product $M_2 \times M'$ can be smoothable; and proves Corollary 2.

REFERENCES

[1]. ADAMS, J. F., Vector fields on spheres. *Ann. Math.*, 75 (1962), 603–632.

[2]. ATIYAH, M. & HIRZEBRUCH, F., Riemann-Roch theorems for differentiable manifolds. *Bull. Amer. Math. Soc.*, 65 (1959), 276–281.

[3]. —— Vector bundles and homogeneous spaces. *Proc. Symp. Pure Math.*, 3. Differential Geometry, *Amer. Math. Soc.*, (1961), 7–38.

[4]. CURTIS, M. & LASHOF, R., On product and bundle neighborhoods. (To appear.)

[5]. EELLS, J. & KUIPER, N., Manifolds which are like projective planes. *Publ. Math. (I.H.E.S.), Paris*, 1962.

[6]. KERVAIRE, M., A manifold which does not admit any differentiable structure. *Comment. Math. Helv.*, 34 (1960), 257–270.

[7]. KERVAIRE, M. & MILNOR, J., Groups of homotopy spheres. (To appear.)

[8]. MILNOR, J., On manifolds homeomorphic to the 7-sphere. *Ann. Math.*, 64 (1956), 399–405.

[9]. —— Differentiable structures on spheres. *Amer. J. Math.*, 81 (1959), 962–972.

[10]. —— Sommes de variétés différentiables et structures différentiables des sphères. *Bull. Soc. Math. France*, 87 (1959), 439–444.

[11]. MUNKRES, J., Obstructions to the smoothing of piecewise-differentiable homeomorphisms. *Ann. Math.*, 72 (1960), 521–554.

[12]. —— Obstructions to imposing differentiable structures. (To appear.)

[13]. SHIMADA, N., Differentiable structures on the 15-sphere and Pontrjagin classes of certain manifolds. *Nagoya Math. J.*, 12 (1957), 59–69.

[14]. SMALE, S., Generalized Poincaré's conjecture in dimensions greater than four. *Ann. Math.*, 74 (1961), 391–406.

[15]. —— On the structure of manifolds. *Amer. J. Math.*, 84 (1962), 387–399.

[16]. STALLINGS, J., The piecewise-linear structure of Euclidean space. *Proc. Cambridge Philos. Soc.*, 58 (1962), 481–488.

[17]. TAMURA, I., Homeomorphy classification of total spaces of sphere bundles over spheres. *J. Math. Soc. Japan*, 10 (1958), 29–43.

[18]. —— 8-manifolds admitting no differentiable structure. *J. Math. Soc. Japan*, 13 (1961), 377–382.

[19]. THOM, R., Les structures différentiables des boules et des sphères. *Colloque Géom. Diff. Globale*. Bruxelles, 1958, 27–35. Centre Belge Rech. Math., Louvain, 1959.

[20]. —— Des variétés triangulées aux variétés différentiables. *Proc. Internat. Congress Math. Edinburgh*, 1958, 248–255. Cambridge Univ. Press, 1960.

[21]. WALL, C. T. C., Classification of $(n-1)$-connected $2n$-manifolds. *Ann. Math.*, 75 (1962), 163–189.

[22]. WHITEHEAD, J. H. C., Manifolds with transverse fields in euclidean space. *Ann. Math.*, 73 (1961), 154–212.

GEOMETRICAL TOPOLOGY

By M. H. A. NEWMAN

In the remarkable series of successes that have been achieved in the last few years in the domain of geometrical topology, the methods of combinatorial and differential topology have both played a part. In this lecture I shall speak of the combinatorial side of the picture, shewing how methods which originated many years ago have been revived and powerfully extended to yield such results as a proof of the generalized Poincaré conjecture (for all but two still doubtful dimensions), and general theorems on embedding and knotting.

The grand divisions of topology are associated with methods of classifying spaces. The most fundamental classifying relation in topology is, of course, homeomorphism, which I shall denote by $X \doteq Y$. Effective classification by homeomorphism alone has long been known to be too hard a problem, and it has been softened in various ways, for example by weakening the classifying relation to homotopy equivalence. Combinatorial and differential topology proceed in the opposite direction, putting in more structure and requiring this also to be preserved. To clarify this statement in so far as it applies to combinatorial topology I must give a few definitions.

We shall be concerned only with *euclidean complexes*, which are composed of rectilinear closed simplexes contained in a euclidean space of some conveniently high dimension. The simplexes form a locally finite set and fit together neatly, in the sense that any two intersect, if at all, in a common face (sub-simplex). Thus a complex is a set of simplexes; but I shall take the usual liberty of using the same word, "complex", for the underlying space $|K|$, which is the point-set union of the simplexes.

Combinatorial equivalence of complexes, which I shall denote by $K_1 \underset{c}{=} K_2$, can be defined in two ways. The simplest is that there should exist a third complex, K_3, obtainable from either K_1 or K_2 by a rectilinear sub-division. This means that each simplex, σ, is replaced by a euclidean complex with the same locus, the subdivided simplexes fitting together at common boundaries in an obvious way.

A different form of definition of combinatorial equivalence, more in the spirit of combinatorial methods, is that K_2 should be derivable from K_1 by a series of elementary transformations, or *moves*, each affecting only a small bunch of simplexes. One may take as moves the insertion of a new vertex in a 1-simplex, σ^1, of K, with consequent division of the simplexes of K containing σ^1; the reverse transformation; and the replacement of K by an isomorphic complex. The relation $K_1 \underset{c}{=} K_2$ defined by these moves is identical with that based on rectilinear subdivision.

A series of combinatorial notions can now be introduced which exactly parallel those of point-set topology. To the topological n-cell and n-sphere, defined to be *homeomorphs* of the n-disc and n-sphere, respectively, correspond the *combinatorial n-cell*, Δ^n, and *combinatorial n-sphere*, Σ^n, which are by definition complexes *combinatorially equivalent* to the n-simplex, σ^n,

Fig. 1. Combinatorial equivalence.

and the simplex-boundary, $\dot{\sigma}^{n+1}$, respectively. A *topological* n-manifold is a separable Hausdorff space in which every point has an open neighbourhood homeomorphic to euclidean space, R^n; in a topological n-manifold with boundary open neighbourhoods homeomorphic to a closed half-space are also allowed. In a *complex* the closed star, $\overline{\mathrm{st}}(a)$, of a vertex a is the set of all simplexes containing a, and their faces; and the link, $\mathrm{lk}(a)$, is the set of simplexes of $\overline{\mathrm{st}}(a)$ not containing a.

A *combinatorial n-manifold* ("an M^n") is a complex in which the link of every vertex is a Σ^{n-1}; K is a combinatorial-manifold-with-boundary if every $\mathrm{lk}(a)$ is a Σ^{n-1} or Δ^{n-1}. The $(n-1)$-simplexes that belong to just one n-simplex form (with their faces) the *boundary*, $\dot{M}^n$, itself an M^{n-1} (without boundary).

The analogue, in this theory, of a continuous map is in many contexts a *piecewise-linear map*, $f: K_1 \to K_2$, which may be thought of as a simplicial map, not of K_1 into K_2 but of K_{1*} into K_{2*}, these being rectilinear subdivisions of K_1 and K_2. An object of the theory is then a space with a class of combinatorially equivalent triangulations rather than a single complex. A systematic treatment on these lines which shews that we have to do with a genuine category, has recently been given by Zeeman [32], but I will continue here to speak of individual complexes.

Equivalence definitions of a combinatorial nature were introduced early in the century by Dehn and Heegaard [9, p. 161]. Theories of the kind I have just described were developed in the 1920's and 30's: in 1930 by Alexander [3] and by myself [16, 17] in 1926, and some initial successes were obtained with the theory. The theorem that I have already mentioned, equating the two definitions of combinatorial equivalence was proved by means of a superposition theorem [18], and some theorems on adding n-cells to n-manifolds, which I will state here since they are closely related to modern work.

Suppose that M^n, Δ^n are a combinatorial manifold-with-boundary and cell, which intersect in a Δ^{n-1} lying in the boundaries of both: i.e. $M^n \cap \Delta^n = \Delta^{n-1} \subseteq \dot{M}^n \cap \dot{\Delta}^n$. *Then*

$$M^n \cup \Delta^n \underset{c}{=} M^n \text{ [16, p. 635, Theorem 8a].}$$

Another theorem also proved in 1926 was that *any two combinatorial n-cells, Δ_1^n, Δ_2^n in the interior of an M^n are similarly situated*, in the sense that there is a piecewise linear equivalence of M^n on to itself carrying Δ_1^n into Δ_2^n (17, Theorem 3]. Indeed there is an "ambient isotopy" of Δ_1^n into Δ_2^n, that is a series of equivalences, $M^n \to M^n$, each moving only a small region of M^n.

What is interesting about these theorems is that, obvious as they may appear, their point-set analogues are all false. The proof that this is so goes back to an example of Antoine [4] in 1920, but Alexander's famous horned-

sphere construction of 1924 [2] is now more familiar. It shews that the upper hemisphere, E^3_+, of the 3-sphere S^3 can be embedded in S^3, by h, so that $S^3 - h(E^3_+)$ is not an open 3-cell. Therefore h cannot be extended to a homeomorphism of S^3 on to itself, i.e. E^3_+ and $h(E^3_+)$ are not similarly situated.

In view of this, the point-set analogue of the $(M^n \cup \Delta^n)$-theorem can also be expected to fail, since in the point-set case $\dot{\Delta}^n - \Delta^{n-1}$ may not be an $(n-1)$-cell, therefore $(M^n \cup \Delta^n)^{\cdot}$ may not be homeomorphic to $\dot{M}^n$, nor *a fortiori* $M^n \cup \Delta^n$ to M^n.

After these and a few other successes, the combinatorial theories ran into obstacles, in the shape of crucial theorems which could be neither proved nor disproved, so formidable that for many years very little progress was made.

The most famous of these intractable problems was the so-called *Hauptvermutung* of combinatorial topology. The purpose of developing combinatorial theories has always been to apply them to the investigation of the underlying spaces, rather than as an end in themselves. The relation $K_1 \underset{c}{=} K_2$ is intended to represent as closely as possible the relation $|K_1| \doteq |K_2|$ between the loci. That $K_1 \underset{c}{=} K_2$ implies $|K_1| \doteq |K_2|$ is almost obvious. The reverse implication, that any two triangulations of a polyhedron are combinatorially equivalent, is the Hauptvermutung. For 2-complexes this was proved in 1943 by Papakyriakopoulos [21]; for 3-manifolds by Moise [14] who fully cleared up the 3-dimensional case. For spaces with a differentiable structure Cairns [7] proved triangulability in 1937, and Whitehead [30] in 1940 the Hauptvermutung within a class of diffeomorphic triangulations. After this nothing more was known until two years ago, when Milnor [13] gave a counter-example for the general Hauptvermutung, in the shape of a 6-dimensional complex with two combinatorially inequivalent triangulations. But this complex is not a manifold: for manifolds the question is still open. However Smale [23, 24] has proved a somewhat weaker result for the n-ball, D^n and n-sphere, S^n. *If* $n \neq 4$, 5(1) *a triangulation of* S^n *or* D^n *known to be a combinatorial* M^n (*or* M^n*-with-boundary*) *is a combinatorial n-sphere or n-cell.* Quite recently Stallings [26] has proved the corresponding theorem for open euclidean space, with $n \neq 4$. These are deep results, but they still leave open the possibility of inequivalent triangulations which are not combinatorial manifolds.

Another problem that proved quite intractable for more than thirty years was the proof or disproof of a theorem somewhat analogous to those on $M^n \cup \Delta^n$ mentioned above. Suppose a Σ^{n-1} embedded as a sub-complex in Σ^n. Are the two residual complexes, that have Σ^{n-1} as a common boundary, combinatorial n-cells? The point-set analogue is again false for $n=3$, as Alexander's horned sphere [2] shews; but he also shewed in the same year, 1924, that when $n=3$ the combinatorial theorem is true [1]. For $n>3$ the question remains unanswered to this day.

However in 1959, B. Mazur [11] gave a remarkably simple proof of a point-set theorem which goes a long way towards an answer, by giving a sufficient restriction on a topological $(n-1)$-sphere $h(S^{n-1})$ in S^n to ensure that the closed complements are n-cells. The condition is that $h(S^{n-1})$ wears

(1) The case $n=7$ (listed as an exception in [24]) was communicated to me verbally by Dr. Smale before this lecture.

a collar, or a *bi-collar*, as Morton Brown now prefers to call it. The theorem is then:

If J is the real interval $[-1,1]$, if $h': S^{n-1} \times J \to S^n$ is an embedding, and if $h(x) = h'(x \times 0)$ for $x \in S^{n-1}$, then $h(S^{n-1})$ splits S^n into two closed n-cells. Mazur's original proof required a further local smoothing condition, but Morton Brown's proof [5], published soon after, dispensed with it. It was also shewn directly by Marston Morse [15] that the extra condition could be removed from Mazur's proof.

If now any Σ^{n-1} is given in a Σ^n, it can be shewn by old-style combinatorial arguments (see [19]) that $|\Sigma^{n-1}|$ is bi-collared in $|\Sigma^n|$; this follows also from Morton Brown's recent result [6], that the existence of a *local* bi-collaring of $h(S^{n-1})$, in a self-explanatory sense, implies that it is globally bi-collared. Therefore $|\Sigma^{n-1}|$ bounds two *topological* n-cells in $|\Sigma^n|$. If the Hauptvermutung for n-cells, in the strong sense that *every* simplicial subdivision of D^n is a combinatorial n-cell, were available, the full theorem, that Σ^{n-1} bounds two Δ^n's, would follow. In the meantime one may try to get a more satisfactory theorem by weakening the combinatorial conditions on the embedded sphere, so that the corresponding conditions can be shewn to hold also for the complementary complexes, a situation in which inductive arguments from dimension to dimension are valid. Such conditions were given in my paper [19]. Let us define(1) a Σ^n and a Δ^n inductively, to be euclidean complexes whose loci are a topological n-sphere and n-cell respectively, and in which the link of each vertex is a Σ^{n-1} or, in the second case, possibly a Δ^{n-1}; the induction starting with Σ^0 = point-pair, Δ^0 = a single point. It then still follows that a Σ^{n-1} in a Σ^n is bi-collared, and therefore bounds two topological n-cells; and this time the induction goes through, to give us the result that, considered as subcomplexes of Σ^n, the two cells are Δ^n's.

I will mention a further simple application of the Mazur-Brown theorem, since it was an essential part of the Stallings-Zeeman combinatorial proof of the Poincaré conjecture for $n \geqslant 5$. If two topological n-cells have their boundaries and nothing else in common it is a simple theorem that their union is an n-sphere. But suppose they overlap: what then? The Mazur-Brown theorem shews that the union is still an n-sphere: *if two open topological n-cells have as their union a compact space X, of which they are themselves open subsets, then $X \doteq S^n$.* Indeed, if the cells are A and $B = \phi(D^n)$, and D^n_α is a smaller ball concentric with D^n, the topological sphere $\phi(\dot{D}^n_\alpha)$ is bi-collared in A, the collar consisting of neighbouring spheres. Therefore $\phi(\dot{D}^n_\alpha)$ bounds a closed n-cell A_0 in A; and it obviously bounds the cell $B_\alpha = \phi(D^n_\alpha)$ in B. Because A and B are open in M, $A_0 \cap B_\alpha = \phi(\dot{D}^n_\alpha)$ and $A_0 \cup B_\alpha = X$. Therefore X is a sphere by the elementary theorem.

I turn now to another important ingredient of much modern work, which will take us back into purely combinatorial theories. In the 1939 paper of J. H. C. Whitehead [29], in which he completely solved the problem of characterising homotopy-equivalence in terms of combinatorial moves, he proved a theorem about what he called "regular neighbourhoods", a term which I now wish to define.

A complex K *collapses* into a subcomplex L of itself if it can be reduced

(1) In [19] called an *n-star-sphere* and *n-star-element*.

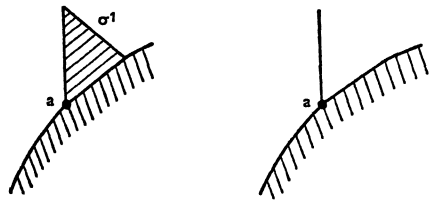


Fig. 2. Collapsing.

to L by a series of moves, $K_i \to K_{i+1}$, in each of which a k-simplex, σ^k, if K_i is removed, having one $(k-1)$-face *free*, (i.e. contained in no other simplex of K_i), while the rest of σ^k is in K_{i+1}.
(The dimension, k, of σ^k may vary from one move to another.) For many applications it is necessary to allow a preliminary subdivision, and I write $K \searrow L$ to mean that K^* collapses to L^*, where K^* is a subdivision of K, and L^* the corresponding subdivision of L.

If M^n is a combinatorial manifold, and $L \subseteq M^n$, *a regular neighbourhood*, N, of L is a sub-complex, N, of M^n such that (1) N is itself an n-manifold with boundary, (2) $N \searrow L$, (3) $L \subseteq \operatorname{int} N$. If M^* is a suitable subdivision of M, and L^* the corresponding subdivision of L, it is not difficult to shew that $\overline{\operatorname{st}(L^*, M^*)}$ is a regular neighbourhood of L. Thus every subcomplex has a regular neighbourhood, if a preliminary subdivision is allowed. But there are many other regular neighbourhoods besides this "thickening". Whitehead's main theorem was:

(1) *All regular neighbourhoods of a finite subcomplex, L, of M^n are combinatorially equivalent*, a theorem proved in true combinatorial style by induction on the number of moves required to collapse N into L.

Immediate corollaries of (1) are:

(2) *If $K \searrow L$ then $N_1(K) \underset{c}{=} N_2(L)$, where N_1, N_2 are arbitrary regular neighbourhoods*; and in particular

(3) *If $K \searrow a$ then $N(K)$ is a Δ^n.*

These theorems are the prototypes, and in some cases the actual basis, of the proofs of many theorems proved in the last two or three years, which tell us that certain fine distinctions between spaces are obliterated by certain dimension-raising processes, such as taking simplicial neighbourhoods ("thickening"), or taking the topological product with a cell or open euclidean space. Whitehead himself proved the first results of this kind about products in 1940, but they have been overtaken by the Poincaré Hypothesis or by modern improvements. But recently Macmillan and Zeeman have shewn [10] that if M^n is a contractible open manifold then $M^n \times R^2 \underset{c}{=} R^{n+2} = R^n \times R^2$; and Stallings [27] has strengthened this to $M^p \times M^q \underset{c}{=} R^{p+q}$ if M^p, M^q are contractible open manifolds with $p+q \geqslant 5$ and $p,q \geqslant 1$. (An *open* manifold has an infinite number of simplexes and no boundary. It is *contractible* if every subcomplex is homotopic to a point.) That "$M^n \underset{c}{=} R^n$" need not be true under the above conditions is known from Whitehead's 1935 example [28] of a contractible 3-space which is not combinatorially, nor in fact topologically, an R^3 [20]. Mazur has also given an interesting example of this kind [12] and Dr. Poenaru will speak in his half-hour talk of analogous results in differential topology.

In a great deal of this recent work, and especially in the combinatorial

proof of the Poincaré conjecture, lemmas have been used by Stallings and Zeeman which are very much in the vein of the Whitehead regular-neighbourhood theorem. Of the rather strenuous proofs of these lemmas I shall not say anything here. Dr. Stallings is to speak on knotting theorems in his half-hour talk and will perhaps give some examples of them. But I would like to indicate briefly how these methods make their appearance in the contemporary scene. (From now on all complexes are finite.)

Zeeman has proved the following lemma [22, 32, 33]:

I. *If L_1^p is a subcomplex of M^n, a combinatorial manifold, and if $p \leqslant n-3$ and $L_1^p \simeq 0$, there exist L_2^q and K in M^n such that $q<p$* and $L_1 \subseteq K \searrow L_2$.

A complex is *p-connected* if every subcomplex of dimension $\leqslant p$ is homotopic to a point. Repeated use of I obviously gives

II. *If M^n is p-connected, and $p \leqslant n-3$, then, given $L_1 = L_1^p \subseteq M_n$, there exists a sequence*

$$L_1 \subseteq K_1 \searrow L_2 \subseteq K_2 \ldots \searrow L_k = \textit{a point.}$$

Let "K is ringed" stand for the statement that K is contained in the interior of a combinatorial n-cell in M^n: $K \subseteq \operatorname{int}\Delta^n \subseteq M^n$. If the following lemma, closely related to Whitehead's regular neighbourhood theorem:

III. *If $K \searrow L$ in M^n, and L is ringed, so is K*; is combined with II there follows the important theorem

IV. *If M^n is p-connected, where $p \leqslant n-3$, then every subcomplex L^p of M^n is ringed.*

For Lemma II is applicable, with $L^p = L_1^p$. If L_k is as in that Lemma, it is ringed, since it is in the interior of its own closed star, a Δ^n. If we assume, using a downward induction on L_r, that L_2 is ringed, then so is K by III, and therefore also $L_1 = L^p$.

It was by using a result of Penrose, Whitehead and Zeeman [22] similar to IV, but somewhat weaker, that Stallings proved the combinatorial form of the Poincaré conjecture for $n \geqslant 7$; and it was by IV itself that Zeeman extended it to $n \geqslant 5$. I would like to indicate how some of the results that *I* have mentioned fitted together to give this proof.

The Poincaré conjecture, in its most general form, said that if a closed topological n-manifold ("closed" means compact and without boundary) is $(n-1)$-connected then it is a topological n-sphere. (In this context "$(n-1)$-connected" means that every map of a k-sphere is homotopic to a point, for $k \leqslant n-1$.) The conjecture was proved by Smale in 1960 [23] for $n \geqslant 5$, with the additional assumption that this manifold carries a differentiable structure; and by Stallings shortly afterwards [26] for $n \geqslant 7$, assuming that it carries a combinatorial-manifold structure [26]. Then Smale shewed [24, 25] that his methods give the combinatorial result for $n \geqslant 5$, and Zeeman [31] that Stallings' methods could also be extended to give this result.

Since I wish only to give an indication of the method I will consider only the case where n is odd, say $n = 2m+1$. Stallings' device was to consider T^m, the *m-skeleton* of M^r, that is the set of simplexes of dimension $\leqslant m$; and T_*^m, the m-skeleton of the dual subdivision. Since $n \leqslant 5$, $m \leqslant n-3$. Therefore, by Lemma III, $T^m \subseteq \operatorname{int} \Delta_1^n$, and $T_*^m \subseteq \operatorname{int} \Delta_2^n$, where Δ_1^n, Δ_2^n are n-cells. Now M^n is a subcomplex of the join(1) of the complexes T'^m and T'^m_*, where the

(1) The *join*, $\sigma\tau$, of two distinct simplexes, σ and τ, is the simplex determined by all their vertices; and the join of two disjoint complexes K and L, is the set of simplexes $\sigma\tau$, where $\sigma \in M$ and $\tau \in L$.

dashes denote *barycentric subdivision*. For indeed typical m-simplexes of T'^m and T'^m_* are $a_0a_1 \dots a_m$ and $a_{m+1} \dots a_n$ respectively, where a_r is the centre of an "old" r-simplex; and a typical n-simplex of M'^n is $a_0a_1 \dots a_n$.

It follows that Δ_2^n can be stretched out along the joining linear segments into a cell Δ_3^n in M^n, whose boundary lies in the neighbourhood int Δ_1^n of T^m.

This is the situation required by the corollary of the Mazur-Brown theorem, mentioned above: int Δ_1^n and int Δ_3^n are open topological n-cells whose union is the space M^n, of which they are open subsets. Therefore M^n is a topological n-sphere.

The missing dimensions, 3 and 4, for which the truth of the conjecture is still an open question, again operate to prevent an inductive proof of a very strong form of the theorem. If one had both the combinatorial Poincaré conjecture and the Smale "weak Hauptvermutung for S^n" for all dimensions it could be proved that if a closed topological n-manifold is $(n-1)$-connected and admits a triangulation, then this triangulation is a combinatorial n-sphere.

I have spoken exclusively of geometrical, and mostly of combinatorial methods, but it will be clear from remarks dropped here and there that differential and combinatorial topology are running curiously parallel and in a sense competing courses at the present time. The differential methods have the advantage of being easily brought into touch with the resources of algebraic topology, but the power of inductive arguments applied to sequences of local modifications is so great that there will always, I believe, remain some problems which can be successfully attacked only by the primitive but penetrating tools of combinatorial topology. But here I am entering upon matters of opinion, and even of taste, a sign that I should bring my remarks to a close.

REFERENCES

[1]. Alexander, J. W., On the subdivision of a 3-space by a polyhedron. *Proc.Nat.Acad.Sci. U.S.A.*, 10 (1924), 6–8.

[2]. —— An example of a simply connected surface bounding a surface which is not simply connected. *Ibid.*, 8–10.

[3]. —— The combinatorial theory of complexes. *Ann. Math.*, 31 (1930), 292–320.

[4]. Antoine, L., Sur la possibilité d'étendre l'homéomorphie de deux figures à leur voisinage. *C. R. Acad. Sci. Paris*, 171 (1920), 661–663.

[5]. Brown, M., A proof of the generalized Schoenflies theorem. *Bull. Amer. Math. Soc.*, 66 (1960), 74–76.

[6]. —— Locally flat embeddings of topological manifolds. *Ann. Math.*, 75 (1962), 331–341.

[7]. Cairns, S. S., Triangulation of the manifold of class one. *Bull. Amer. Math. Soc.*, 41 (1935), 549–552.

[8]. Curtis, M. L. & Zeeman, E. C., On the polyhedral Schoenflies theorem. *Proc. Amer. Math. Soc.*, 11 (1960), 888–889.

[9]. Dehn, M. & Heegaard, P., Analysis Situs. *Enz. d. Math. Wiss.*, III AB3 (1907).

[10]. McMillan, D. R. & Zeeman, E. C., On contractible open manifolds. *Proc. Cambridge Phil. Soc.*, 58 (1962), 221–224.

[11]. MAZUR, B., On imbeddings of spheres. *Bull. Amer. Math. Soc.*, 65 (1959), 59–65.

[12]. —— A note on some contractible 4-manifolds. *Ann. Math.*, 73 (1961), 221–228.

[13]. MILNOR, J., Two complexes which are homeomorphic but combinatorially distinct. *Ann. Math.*, 74 (1961), 575–590.

[14]. MOÏSE, E. E., Affine structures in 3-manifolds. (V). *Ann. Math.*, 56 (1952), 96–114.

[15]. MORSE, M., A reduction of the Schoenflies extension problem. *Bull. Amer. Math. Soc.*, 66 (1960), 113–115.

[16]. NEWMAN, M. H. A., On the foundations of Combinatory Analysis Situs. *Proc. Acad. Sci. Amsterdam*, 29 (1926), 611–641.

[17]. —— On the superposition of n-dimensional manifolds, *J. Lond. Math. Soc.*, 2 (1926), 56–64.

[18]. —— Combinatory topology and euclidean n-space. *Proc. Lond. Math. Soc.*, Ser. 2, 30 (1930), 339–346.

[19]. —— On the division of euclidean n-space by topological $(n-1)$-spheres, *Proc. Roy. Soc. A.*, 257 (1960), 1–12.

[20]. NEWMAN, M. H. A. & WHITEHEAD, J. H. C., On the group of a certain linkage. *Quart. J. Math.*, Ser. 2, 8 (1937), 14–21.

[21]. PAPAKYRIAKOPOULOS, CN., A new proof of the invariance of the homology groups of a complex. *Bull. Math. Soc. Grèce*, 22 (1943), 1–154.

[22]. PENROSE, R., WHITEHEAD, J. H. C. & ZEEMAN, E. C., Imbeddings of manifolds in euclidean space. *Ann. Math.* 73 (1961), 613–623.

[23]. SMALE, S., The generalised Poincaré conjecture in higher dimensions. *Bull. Amer. Math. Soc.*, 3, 66 (1960), 373–375.

[24]. —— The generalised Poincaré conjecture in dimensions greater than 4. *Ann. Math.*, 74 (1961), 391–406.

[25]. —— Differentiable and combinatorial structures on manifolds. *Ann. Math.*, 74 (1961), 498–502.

[26]. STALLINGS, J. R., Polyhedral homotopy-spheres. *Bull. Amer. Math. Soc.*, 60 (1960), 485–488.

[27]. —— The piecewise linear structure of euclidean space. *Proc. Cambridge Philos. Soc.*, 58 (1962), 481–489.

[28]. WHITEHEAD, J. H. C., A certain open manifold whose group is unity. *Quart. J. Math.*, Ser. 2., 6 (1935), 268–279.

[29]. —— Simplicial spaces, nuclei and m-groups. *Proc. Lond. Math. Soc.*, Ser. 2, 45 (1939), 243–327.

[30]. —— On C^1-complexes. *Ann. Math.*, 41 (1940), 825–832.

[31]. ZEEMAN, E. C., The generalised Poincaré conjecture. *Bull. Amer. Math. Soc.*, 67 (1961), 270.

[32]. —— On polyhedral n-manifolds I and II. *Topology of 3-manifolds*, (ed. M. K. Fort, Jr.), Prentice-Hall, 1962.

[33]. —— On polyhedral manifolds. (To appear.)

SOME ASPECTS OF LINEAR AND NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

By L. NIRENBERG[1]

The speaker's committee has asked me to speak on some of the recent work in nonlinear partial differential equations. I would like to start by quoting from Prof. G. Temple's address [58] at the Edinburgh Congress—"the closely guarded secret of this subject (differential equations) is that it has not yet attained the status and dignity of a science but still enjoys the freedom and freshness of pre-scientific study.... The work of classification and systemization of specimens has hardly begun." Indeed most of the specimens considered still deal directly with equations arising from physical theory or from geometric problems, and the work seems rather special and indeed peculiar to some mathematicians from other fields. In the past decade, however, the general theory of linear partial differential equations with constant coefficients has been greatly developed; this progress is described elsewhere in this Congress.

In this talk which is necessarily limited by my own restricted knowledge and inclinations I shall not deal with equations arising from physical problems. Thus, in particular, and also because of time restrictions, I will not describe the intensive attack that has been made again in recent years, after the basic work by J. Leray in the 30's, on the existence problems for the Navier-Stokes equations of fluid mechanics, except to make reference to the recent book by O. A. Ladyzhenskaya [28] (soon to appear in English translation) for a clear exposition of much of the recent developments, as well as expository articles by G. Prodi [52], J. Serrin [55] and R. Finn [10] where further references may be found. (Rather than have a lengthy bibliography, I have tried to limit it to papers containing fairly comprehensive lists of references.)

I wish to describe some recent developments in existence and regularity theory for nonlinear boundary value problems, mainly for elliptic differential equations, with side remarks for parabolic and hyperbolic equations. Since this lecture is directed to nonexperts, I would like to give some indication of the techniques used and not merely list the latest and strongest results which are often quite complicated to formulate.

§ 1. Most results for nonlinear problems are still obtained via linear ones, i.e. *despite* the fact that the problems are nonlinear not *because* of it. So we shall begin with the simplest question for a nonlinear problem, a perturbation problem, in which the problem differs slightly from a linear one; this simply involves the implicit function theorem in a suitable framework.

To fix notation we usually treat functions $u(x)$, $x=(x_1,\ldots,x_n)$ defined in a bounded domain in E^n with smooth boundary; u may represent a system of functions. Differentiation is denoted by $D=(D_1,\ldots,D_n), D_j=\partial/\partial x_j$,

[1] The author is a Sloan Fellow.

$D^\alpha = D_1^{\alpha_1} \dots D_n^{\alpha_n}$ for $\alpha = (\alpha_1, \dots \alpha_n)$ with the α_j integers $\geqslant 0$, and D^α has order $|\alpha| = \sum \alpha_j$. We also write the first and second derivatives of u as $u_{x_i}, u_{x_i x_j}$. For $k \geqslant 0$ an integer we use the norm

$$|u|_k = \sum_{|\alpha| \leqslant k} \sup |D^\alpha u|$$

for the space C^k.

Consider a nonlinear partial differential equation of order m depending on some parameters ε

$$F(\varepsilon, x, u, D^\alpha u) = 0, \tag{1.1}$$

we are interested in solutions satisfying, say, homogeneous conditions on the boundary (or part of the boundary) assuming that u_0 is a solution for $\varepsilon = 0$. If $L_0 = L(u_0)$ is the linearized operator about the solution u_0, the first variation of F:

$$L_0 v = L(u_0) v = \sum_\alpha \frac{\partial F}{\partial D^\alpha u}(0, x, D^\alpha u_0) D^\alpha v,$$

then one writes the nonlinear equation (by writing F as a linear term in $(u - u_0)$ plus a remainder term of higher order)

$$L_0(u - u_0) = R[u, \varepsilon],$$

where R involves higher order terms in $u - u_0$ and its derivatives. If the linearized problem

$$L_0 v = f$$

with the homogeneous boundary conditions has a solution (say unique) $v = L_0^{-1} f$, such that if f belongs to a certain class of functions then the mth order derivatives of the solution v belong to this same class, the Picard iteration:

$$u_{n+1} = L_0^{-1} R[u_n, \varepsilon]$$

will yield a solution. This is just the usual implicit function theorem. In applying it the main thing to verify is the property of the solution $L_0^{-1} f$—that it is sufficiently differentiable. Otherwise one loses differentiability in each iteration step.

For elliptic partial differential operators L_0 and for a wide class of boundary conditions it has been shown that L_0^{-1} has the right properties, in a suitable function space. This is a consequence of the, so-called, Schauder type estimates for such equations; see S. Agmon, A. Douglis, L. Nirenberg [1] (these estimates have been derived also for general elliptic systems in part 2 soon to appear), see also F. Browder [3].

We remark that the class of boundary conditions for which such estimates hold are those satisfying certain algebraic relations involving their leading parts and the leading part of the operator L_0. (A simple way of describing these conditions at the boundary point x_0 is to consider the leading part L_0' of L_0, and leading parts of the boundary differential operators with coefficients having their value at x_0, i.e. constant, and to imagine the boundary as flat at x_0, and require that the only bounded finite sum of exponentials v satisfying $L_0' v = 0$ in the half space, and satisfying the leading part

of the boundary conditions on the boundary of the half space, and tending to zero as we go to infinity normal to the boundary, is $v \equiv 0$.) Recently there have occurred some situations in which the lower order terms also play a crucial role, in particular, in the work of C. B. Morrey [39] and its extension by J. J. Kohn [27] on complex analytic manifolds; the general situation here has yet to be cleared up. For planar boundaries and constant coefficients, in the case of one operator, L. Hörmander [22] has resolved the problem of regularity at the boundary.

Let us return to the general perturbation problem or implicit function theorem. The Picard iteration method described there converges like a geometric series. Namely $\|u_{n+1}-u_n\| \leqslant \theta \|u_n - u_{n-1}\|, \theta < 1$. Here $\| \ \|$ denotes some norm. It may occur in practice that the operator L_0^{-1} "loses" derivatives. For instance, if we work with the spaces C^k (which is not very suitable for elliptic problems) we want L_0^{-1} to map C^ϱ boundedly into $C^{\varrho+m}$ but it may only map C^ϱ boundedly into $C^{\varrho+m-\sigma}$ (thus a loss of σ derivatives). Then Picard iteration doesn't work. In 1956 J. Nash [46] was able to treat a special situation just of this kind occurring in his work on isometric embedding of Riemannian manifolds. He did this by means of a rather remarkable but complicated and, to me, mysterious scheme involving a combination of approximation and "smoothing" of functions. This method was set into a general framework by J. Schwartz [54] who proved a general implicit function theorem. Recently J. Moser [44] has succeeded in giving a clear and conceptually straightforward proof of this and I would like to describe the idea of this important method—sticking still to a partial differential equation although the method works for general nonlinear functional equations. It involves two devices:

(1) An iteration scheme—Newton's instead of Picard's. Newton's method is the following

$$u_{n+1} - u_n = -L(u_n)^{-1} F(\varepsilon, x, D^\alpha u_n) = -L(u_n)^{-1} F[u_n].$$

Thus we assume not only that $L(u_0)$ is invertible but also $L(u)$, for u sufficiently "close" to u_0. By the mean value theorem we have

$$|F[u_n] \underbrace{- F[u_{n-1}] - L(u_{n-1})(u_n - u_{n-1})}_{0}|_\varrho \leqslant \text{const}\, |u_n - u_{n-1}|^2_{m+p}. \tag{1.2}$$

If $L(u_n)^{-1}$ were a bounded linear map of $C^{\varrho+m}$ into C^ϱ we would consequently have

$$|u_{n+1} - u_n|_{m+\varrho} \leqslant \text{const}\, |u_n - u_{n-1}|^2_{m+\varrho}$$

and hence very rapid convergence of the iteration scheme.

Moser uses this scheme together with a smoothing device (also used by Nash), an operator T_N depending on a parameter N:

(2) For every N large T_N is a linear mapping of functions into C^∞ functions with the properties (here $\delta \geqslant 0$ is a fixed number):

For all integers $p, q \geqslant 0$

(i) $|T_N u|_{p+q} \leqslant \text{const}\ N^{q+\delta} |u|_p$,

(ii) $|u - T_N u|_p \leqslant \text{const}\ N^{-q+\delta} |u|_{p+q}$.

The constants depend on p and q. Condition ii) expresses the fact that the approximation $T_N u$ is very close to u if u itself is very smooth. (In case we

consider functions u defined in all E^n [or periodic in all variables] such smoothing operators may be obtained by multiplying the Fourier transform of u [or Fourier coefficients of u] by functions with compact support, or, what is equivalent, by convoluting u with *suitable* kernels $K(Nx)$ such that $\int K(x)dx=1$. One even obtains (i) and (ii) with $\delta=0$.)

Let us see how to obtain a solution u of (1) in some class $C^{m+\varrho}$ assuming that the inverse of the linear operator $L(u)$ is defined for $|u-u_0|_m$ sufficiently small and is a bounded map of C^ϱ into $C^{\varrho+m-\sigma}$, i.e. it loses σ derivatives. Moser's iteration scheme is to choose $N_{n+1}=N_n^{3/2}$, and set

$$u_{n+1}-u_n=-T_{N_{n+1}}L(u_n)^{-1}F[u_n]. \tag{1.3}$$

With $\mu=3(\sigma+\delta), l=35(\sigma+\delta+1)$ one shows inductively that

$$|u_{n+1}-u_n|_{m+\varrho}\leqslant \text{const } N_{n+1}^{-\mu},$$

$$|u_n-u_0|_{m+\varrho+l}\leqslant \text{const } N_n^l.$$

These estimates are not very difficult to prove using (i) and (ii) and assuming that u_0, the initial solution, belongs to $C^{m+\varrho+l}$. The iterates u_n converge to a solution u in $C^{m+\varrho}$; u is thus considerably less smooth than the initial solution u_0. In working with general nonlinear operator equations Moser imposes a third condition (condition (3) in [44]) which is usually satisfied in practice, and which we have omitted here.

This result and its proof are set within the framework of linear spaces. For nonlinear problems this seems slightly unnatural, but there is as yet no systematic theory operating without these spaces.

Moser has applied similar techniques in studying stability of C^k solutions of ordinary differential equations. In this work he is not able to operate in a linear function space. Furthermore the nonlinear character of the equations is used in an essential way, indeed he obtains results *because* of the nonlinearity not despite it. This work is related to the work of Kolmogorov and Arnold on stability for analytic nonlinear differential equations; the use of Newton's method was suggested by that work (see Moser's report at the Congress).

§ 2. We turn to more special equations—second order quasilinear equations of the form (using summation convention)

$$a_{ij}(x, u, Du)\,u_{x_ix_j}=f(x, u, Du) \tag{2.1}$$

a_{ij} positive definite matrix, or equations in divergence from

$$\frac{\partial}{\partial x_i}a_i(x, u, Du)=f(x, u, Du), \tag{2.2}$$

$\partial a_i/\partial u_{x_j}$ positive definite matrix, such as arise from regular variational problems

$$\delta\int F(x, u, Du)\,dx=0, \tag{2.3}$$

for which the Euler equation is

$$\left(\frac{\partial F}{\partial u_{x_i}}\right)_{x_i} = \frac{\partial F}{\partial u}. \tag{2.4}$$

One may also write (2.2) in the "weak" from

$$\int (a_1(x, u, Du)\zeta_{x_i} + f(x, u, Du)\zeta)\,dx = 0 \tag{2.2'}$$

for all $\zeta \in C_0^\infty$, i.e. for all functions $\zeta \in C^\infty$ with compact support in the domain. Since Hilbert formulated the problem of showing regularity of solutions of (2.3) or (2.4) these equations have received much study concerned also with existence theorems for boundary value problems.

Much of the work in connection with these equations has been directed to finding a priori bounds for solutions of the equations, i.e. bounds for a solution and its derivatives, assuming that a smooth solution exists. The search for such bounds goes back to the fundamental work of *S*. Bernstein, and their use has been fully clarified in the basic work of J. Leray and J. Schauder [33].

We can give a brief indication, to those who are not familiar with the field, how these bounds may be used in proving existence of a solution of, say, (2.1) with given boundary values. Assuming that we know how to solve linear elliptic problems (elliptic here means that a_{ij} is positive definite), insert in the coefficients of (2.1) a function v (and its first derivatives), and solve the corresponding linear elliptic equation for a function u taking on the given boundary values. This defines a transformation $u = T[v]$ which, because u belongs to a higher differentiability class than v, can be shown to be a compact operator. If we had a priori bounds for solutions of such linear elliptic equations which involve practically no knowledge of the coefficient, say that u and some derivatives are bounded, then we could assert that T maps the set of functions satisfying these conditions into itself. Because T is compact it would follow, by the Schauder fixed point theorem, that T has a fixed point—which is then a solution of (2.1). Thus we see why it is useful to obtain *a priori* bounds for solutions of linear equations under minimal assumptions on the coefficients.

A more general procedure is to obtain a priori bounds for solutions of the nonlinear problem by using some special features or structure of the equation. Then one connects the equation by a one parameter family of equations $E_t, 0 \leqslant t \leqslant 1$, to an equation E_1, which one can solve. If a priori bounds for solutions of all these equations can be obtained, e.g. that some norm $\|u\|$ of the solutions u remain bounded, $\|u\| \leqslant K$, then in the sphere $\|u\| \leqslant K$ in the Banach space with $\| \ \|$ as norm one may try to use the Leray-Schauder [33] theory of degree of mapping (see also Leray [32] and M. Nagumo [45]) to show that all the equations E_t possess solutions starting from E_1, since the solutions cannot cross the boundary of the sphere $\|u\| \leqslant K$.

Before taking up any special equations let me mention a few tools of the calculus which are used; these sometimes enable one to treat certain lower order terms in an equation as minor perturbation terms. In the following we use $|D^j u|_{L_q}$ to denote the sum of the L_q norms of all derivatives of u of order j, and denote constants by c; also $j < m$ below.

Sobolev inequalities:

$$|D^j u|_{L_q} \leqslant c(|D^m u|_{L_p} + |u|_{L_p})$$

if

$$\frac{1}{q} = \frac{1}{p} - \frac{m-j}{n} > 0.$$

Interpolation inequalities:

$$|D^j u|_{L_q} \leqslant c\,|D^m u|_{L_p}^{a}\,|u|_{L_r}^{1-a} + c\,|u|_{L_r}, \quad \text{for} \quad \frac{j}{m} \leqslant a \leqslant 1$$

if

$$\frac{1}{q} = \frac{j-am}{n} + \frac{a}{p} + \frac{1-a}{r} > 0.$$

The latter inequalities enable one to say that the set of bounded functions with derivatives of order m in L_p form an algebra—a fact that is useful in nonlinear problems. (Proofs of these may be found in L. Nirenberg [48], lecture 2, and in E. Gagliardo [13].) The interpolation inequalities are special cases of general abstract interpolation inequalities that have been found in the last few years which are generalizations of the Riesz-Thorin convexity theorem. These should prove useful in nonlinear problems. We content ourselves here with the following references: E. Gagliardo [14], A. P. Calderon [4], J. L. Lions [34] where further references may be found. Some applications to nonlinear problems are indicated in chapters 4, 8 and 10 of J. L. Lions [35].

Another concept that is used in the study of nonlinear differential equations is that of "weak" or "generalized" solution. Using W_p^m, (or $H_{m,p}$) to denote the completion in the norm $\sum_{|\alpha| \leqslant m} |D^\alpha u|_{L_p}$ of C^∞ functions (the preceding inequalities extend to functions lying in these spaces) one says, say, that $u \in W_p^1$ is a weak solution of

$$(a_{ij} u_{x_i})_{x_j} = f$$

if

$$\int (a_{ij} u_{x_i} \zeta_{x_j} + \zeta f)\, dx = 0 \quad \text{for all} \quad \zeta \in C_0^\infty.$$

In seeking to prove existence of regular solutions of nonlinear elliptic equations it is often convenient (and usually simple) to prove first the existence of a weak solution. Then one attempts to show that this solution is regular.

§ 3. In considering second order equations I shall divide the results to be described into two classes; those similar to results for linear equations and others which are more nonlinear—starting with the former. In seeking bounds for solutions of equations (2.1) and (2.2) one has first, in case the maximum principle applies, a bound for the solution itself. In the 1930's Schauder developed an extensive theory for linear equations with Hölder continuous(1) coefficients (this has been extended in [1]). Applying this theory one obtains bounds for all derivatives of a solution of (2.1) if a bound

(1) A function v is Hölder continuous, or satisfies a Hölder condition if

$$\sup \frac{|v(x) - v(y)|}{|x-y|^\alpha} < \infty, \quad \alpha < 1.$$

for its first derivatives and their Hölder continuity is known. So we shall restrict ourselves to these. For two dimensional, $(n=2)$ problems, after Schauder and Leray (see C. Miranda [36] for other references) the basic work in getting bounds for solutions, from which much further work stems is due to C. B. Morrey [37–38]. This work is connected with the theory of quasi-conformal mappings (see also L. Bers, L. Nirenberg [2]). In higher dimensions little was known (see, however, H. O. Cordes [5] where other references may be found) until in 1957 E. de Giorgi [6] (and in 1958 John Nash [47], for parabolic equations, which we mention later) succeeded in estimating the Hölder continuity in compact subsets of solutions u of linear elliptic equations in divergence, i.e. in the weak, form

$$\int a_{ij} u_{x_i} \zeta_{x_j} dx = 0 \quad \text{for all } \zeta \in C_0^\infty, \tag{3.1}$$

assuming the equations to be uniformly elliptic, i.e. to satisfy

$$m \sum_j \xi_j^2 \leqslant a_{ij} \xi_i \xi_j \leqslant M \sum \xi_j^2,$$

for suitable positive constants m, M. They showed that a solution u in W_2^1 is Hölder continuous. De Giorgi used this result to prove the analyticity of solutions of uniformly elliptic variational problems,

$$\delta \int F(Du)\, dx = 0.$$

Any first derivative u_{x_k} of such a solution satisfies an equation of the form (3.1), namely

$$\int F_{u_i u_j} u_{x_k x_i} \zeta_{x_j} dx = 0 \quad \text{for all } \zeta \in C_0^\infty$$

and u_{x_k} is easily seen to belong to W_2^1. Therefore u_{x_k} is Hölder continuous, and the analyticity of u then follows from previously known results. The proof of Hölder continuity of solutions of (3.1) involves two steps:

(*a*) Obtaining a bound for u, of the form

$$|u(x)|^2 \leqslant \frac{\text{const}}{R^n} \int_{|y-x|<R} |u(y)|^2 dy,$$

(*b*) Then deriving a Hölder condition for u.

After de Giorgi simpler proofs have been given by J. Moser [42–43] and G. Stampacchia [56]. Stampacchia's proof of (*a*), which is closer to de Giorgi's, involves choosing for ζ in (3.1) a C_0^∞ function ψ times a truncation of $u, u^k = \max(u-k, 0)$. This choice gives immediately a bound on the L_2 norm of grad u^k. Using Sobolev's inequalities one obtains a bound for the measure σ_k of the set where $u > k$. Repeating this argument again one obtains an estimate for σ_h in terms of σ_k, $h > k$ and one finds that for $h \geqslant$ some $k_0, \sigma_h = 0$, i.e. $u \leqslant h$ almost everywhere. Moser's proof of (*a*) works with the L_p norm of u in place of σ_h; he chooses for ζ a function $\psi \in C_0^\infty$ times a power of u (note that though this is a linear problem nonlinear operations are per-

formed), and shows that the L_p norm of u on a compact subset remains bounded as $p \to \infty$.

Stampacchia (see his Congress lecture and [57] has used his method of proof to prove the following

Maximum principle: Let $u \in W_2^1$ be a weak solution in a domain $\mathfrak{D}$ of

$$(a_{ij}u_{x_i} + f_j)_{x_j} \geqslant 0, a_{ij}\xi_i\xi_j \geqslant m\xi_i^2, \quad u = \phi \quad \text{on boundary} \tag{3.2}$$

i.e.
$$\int (a_{ij}u_{x_i} + f_j)\zeta_{x_j} dx \leqslant 0 \text{ for } \zeta \geqslant 0, \quad \zeta = 0 \quad \text{on boundary.} \tag{3.2'}$$

then
$$u \leqslant \max \phi + \frac{K(\mathfrak{D}, p)}{m} \sum_j |f_j|_{L_p} \quad \text{for} \quad p > n.$$

We may give a brief sketch of the proof. Take for ζ the function $\max(u-k, 0)$ with $k > \max \phi$. Then from (3.2′) we find, letting D_k denote the set where $u > k$, and σ_k its measure

$$m\int_{D_k} |\operatorname{grad} u|^2 dx \leqslant \int_{D_k} |\sum f_j u_{x_j}| dx \leqslant \left(\int_{D_k} |\operatorname{grad} u|^2 dx\right)^{1/2} \left(\int \sum |f_j|^2 dx\right)^{1/2}$$

or, by Hölder's inequality,

$$m\left[\int_{D_k} |\operatorname{grad} u|^2 dx\right]^{1/2} \leqslant (\sum |f_j|_{L_p}) \sigma_k^{(p-2)/2p}.$$

By Sobolev's inequality it follows that for some constant C

$$\left(\int_{D_k} |u-k|^q dx\right)^{1/q} \leqslant \frac{C}{m} \sum |f_j|_{L_p} \sigma_k^{(p-2)/2p}, \quad \text{for} \quad q = \frac{2n}{n-2}.$$

Therefore restricting the integration on the left to D_h, for $h > k$ we find

$$\sigma_h^{(n-2)/2n} \leqslant \frac{\frac{C}{m}\sum |f_j|_{L_p}}{h-k} \sigma_k^{(p-2)/2p} \cdot h > k.$$

For $p > n$ the power of σ_k on the right is greater than that of σ_h on the left and one shows consequently, fairly readily, that

$$\sigma_h = 0 \quad \text{for} \quad h \geqslant \max \phi + \frac{\text{const}}{m} \sum |f_j|_{L_p}$$

which is the desired result.

The most elegant proof of the Hölder continuity (b) has been given by Moser in [43] (his second proof). It is based on a

Harnack inequality: any nonnegative solution u of

$$(a_{ij}u_{x_i})_{x_j} = 0, \quad m\sum \xi_j^2 \leqslant a_{ij}\xi_i\xi_j \leqslant M \sum \xi_j^2$$

in a unit sphere, satisfies in any concentric smaller sphere S

$$\max_S u \leqslant C \min_S u,$$

where C depends only on m, M and S.

This yields easily the fact that if u is a bounded solution of (3.1) in a sphere then its oscillation (max-min) in a concentric sphere of half the radius is bounded by $\theta(<1)$ times its oscillation in the full sphere. Hölder continuity then follows easily in turn from this result.

(Moser's proof makes use of a lemma by F. John and L. Nirenberg [25] which has also been used by F. John in an interesting paper [24] studying (nonlinear) mappings of a domain in E^n into E^n which differ little from an isometry.

LEMMA. *Suppose u is an integrable function in a cube C_0; denote by u_c the average of u in a parallel subcube C. Assume that for every such subcube C the inquality*

$$(|u-u_C|)_C \leqslant 1$$

holds then $u \in L_p$ for avery p, in fact $\int_{C_0} e^{a|u-u_{C_0}|}dx \leqslant const$, for some suitable constant $a>0$.)

The results of de Giorgi have also been extended to hold in the full domain (not just compact subsets) for solutions satisfying various boundary conditions. These extensions yield existence theorems for uniformly elliptic equations (2.2) and (2.4). The results have also been extended to certain nonuniformly elliptic equations. The main extensions of this kind, and applications to boundary value problems, have been given by C. B. Morrey, O. A. Ladyzhenskaya and N. N. Uraltseva, O. A. Oleinik and C. N. Kruzhkov. The equations are not required to be uniformly elliptic but, say for (2.2), the coefficients a_i are permitted to behave like polynomials in the u_{x_i}. The conditions on the coefficients are too complicated to be stated here, but there are examples showing that these conditions are rather natural. I mention here just one condition involved in one case for (2.2), here $V^2=1+|u|^2+|\text{grad } u|^2$, $\xi=(\xi_1,\ldots,\xi_n)$ a real vector;

$$mV^k|\xi|^2 \leqslant \frac{\partial a_i}{\partial u_{x_j}}\xi_i\xi_j \leqslant MV^k|\xi|^2 \quad (k>-1),$$

and a_i is permitted to grow like V^{k+1}. In case (2.2) comes from a variational problem (2.3) the further conditions express the requirement that the functional $\int F(x,u,Du)dx$ be of class C^2 in the space W^1_{k+2}.

This work, together with complete references, may be found in the clear expository article by Ladyzhenskaya and Uraltseva [29], in the papers [40–41] by Morrey and in Oleinik, Kruzhkov [50].

I will just mention one result from the paper [29]. Consider the simple nonlinear equation:

$$\Delta u = \text{quadratic in the } u_{x_i}, u=\phi \text{ on boundary}$$

(here Δ represents the Laplace operator, this is a very special case of a result in [29]). It is known, and in fact not difficult to derive with the aid of the interpolation inequalities mentioned above, that if ϕ is small, or if

the coefficients in the quadratic are small, or even if the oscillation of u in a fixed region is small, then one can obtain a priori estimates for all derivatives of u and prove the existence of a solution. If the quadratic is replaced by something of higher degree in u_{x_i} then a regular solution need not exist, while for lower degree it is quite easy to prove existence. It has been an open question for some time to see if there is existence in the case of a general quadratic. This is settled in the affirmative in [29] by Ladyzhenskaya and Uraltseva. The main step is an estimate of the Hölder continuity of u, and its derivation is similar to their extension of de Giorgi's work, and is therefore not very simple.

G. Stampacchia [57] and D. Gilbarg [16] have also obtained existence theorems with the aid of de Giorgi's result and its extensions. In particular for equations in a strictly convex domain they have independently obtained some very clean results. To mention one, consider a variational problem

$$\delta \int F(u_{x_i})\,dx = 0, \quad u = \phi \text{ (smooth) on boundary,}$$

which is supposed to be regular, i.e. for which the Euler equation is elliptic, $\partial^2 F/\partial u_{x_i}\partial u_{x_j}$ is positive definite—then F is a convex function. They prove that there exists a unique (regular) solution of this problem. The proof is based on the a priori estimate

$$|u| + |\text{grad } u| \leqslant K(\mathfrak{D},\phi)$$

where the constant K depends only on the (strictly convex) domain $\mathfrak{D}$ and the boundary function ϕ, K does not depend on the function F. The estimate is easily derived with the aid of the maximum principle (which holds also for grad u) and a comparison function which is linear. Using the estimate Stampacchia's existence proof proceeds as follows: He modifies the function F for $|\text{grad u}| \geqslant K$ so that the new function $\tilde{F}$ is uniformly regular, i.e. the eigenvalues of the Hessian matrix are bounded from above and below by fixed positive constants; this indeed is the main step of the proof. For the variational problem $\delta\int \tilde{F}(u_{x_i})dx = 0, u = \phi$ on boundary, it is very easy to find a weak solution in W_2^1. Using the extension of de Giorgi's result to the boundary for uniformly elliptic problems it follows that this solution is regular. But it must also satisfy the a priori estimate above, and hence is seen to be a solution of the original problem.

A word about parabolic equations. Using quite different methods J. Nash [47] proved the analogue of de Giorgi's result for parabolic equations in full space of the form

$$\frac{\partial u}{\partial t} = (a_{ij}u_{x_i})_{x_j}.$$

This has been generalized by Oleinik and Kruzhkov [50], and rederived and extended by Ladyzhenskaya and Uraltseva [30] by adapting de Giorgi's methods. The analogue of the Schauder theory for linear equations had been established by A. Friedman [11] who treated also nonlinear problems [12]. Various existence (as well as nonexistence theorems), a priori estimates and stability theorems, have also been given by A. F. Filippov [8] and S. Kaplan [26]. Unfortunately we cannot describe these results here.

What about higher order elliptic equations? Very little is known, but recently M. I. Vishik [59–60] has made an interesting beginning by proving existence of weak solutions for certain classes of quasilinear systems (written in divergence form) whose coefficients behave essentially like polynomials in the unknown functions and their derivatives. The analogue of the de Giorgi results, giving regularity, has yet to be found.

§ 4. I would like to turn now to second order elliptic equations which, are "truly nonlinear", in the sense that one has new phenomena occurring which have no analogue for linear equations. The best known specimen (of which most of the others are variations and generalizations) is the minimal surface equation, expressing the fact that the surface is a solution of the variational problem

$$\delta \int dA = 0,$$

where dA is element of surface area. If the surface has simple projection on the (x,y) plane, i.e admits the representation $u=u(x,y)$ in (x,y,u) space, then u satisfies

$$(1+q^2)u_{xx}-2pqu_{xy}+(1+p^2)u_{yy}=0, \quad p=u_x, \quad q=u_y.$$

Geometrically this asserts that the spherical image mapping of the surface by the unit normal to the unit sphere is conformal. For this equation have been known for some years a variety of results showing marked difference from linear equations. To mention just a few (references can be found in the papers listed later):

1. S. Bernstein's classical theorem that the only solution defined over the whole plane is a linear function.

2. Removability of isolated singularities (L. Bers).

3. It is possible to estimate the Gauss curvature of the surface $u=u(x,y)$ defined in $x^2+y^2<R^2$ at $(0,0)$ in terms of the gradient of u at $(0,0)$. (E. Heinz, E. Hopf.) On letting $k\to\infty$ this estimate yields another proof of Bernstein's result 1.

In recent years these results have been extended in various directions. R. Osserman [51] showed that Bernstein's theorem holds in general for a *complete* (i.e. geodesics can be extended to have infinite length) minimal surface which need not have a simple projection on a whole plane, provided that the spherical image of the surface deletes an open set on the sphere. This was done by showing that such a (simply connected) surface, when regarded as a Riemann surface, is of parabolic type. He also obtained estimates for the Gauss curvature analogous to those in 3.

R. Finn has treated boundary value problems, obtained estimates as well as the removable singularity theorem, and other results, for a class of equations which he calls of minimal surface type (see [9] and other papers by Finn)

$$a(p,q)u_{xx}+2b(p,q)u_{xy}+c(p,q)u_{yy}=0$$

of the form
$$(A(p,q))_x+(B(p,q))_y=0.$$

In this work one uses results for quasi-conformal mappings. Some of his results are new also for minimal surfaces.

I would like to call attention to an interesting paper by H. B. Jenkins [23] in which he treats a class of *nonparametric* variational problems of the form

$$\delta \int F(\text{unit normal})\, dA = 0,$$

where F is homogeneous function of three variables of degree zero such that $o < m \leqslant F \leqslant M$ and the surface $|x| F(x) = 1$, is closed convex with Gauss curvature > 0; $F = 1$ for minimal surfaces. For such variational problems, to which he shows that the theory of Finn applies, he derives a variety of interesting results including Osserman's extensions of Bernstein's result. More recently Jenkins together with J. Serrin have derived a priori estimates for derivatives of solutions of such variational problems, improving Finn's results and, in particular, have derived an interesting *Harnack inequality* for solutions with simple projection on a disc, i.e. solutions of the form

$$u = u(x, y) > 0 \quad \text{in } \textit{unit circle};$$

namely $$\phi_1(r, u(0)) \leqslant u(x, y) \leqslant \phi_2(r, u(0)), \; r = \sqrt{x^2 + y^2},$$

where however $\phi_2(r, u(0)) \to \infty$ *as* $r \to r_0(u(0)) < 1$.

This behavior of ϕ_2 is also shown to be appropriate even for minimal surfaces. Furthermore they show that

$$r_0(t) \to 0 \text{ as } t \to \infty \quad \text{and} \quad r_0(t) \to 1 \quad \text{as} \quad t \to 0.$$

Finn has also recently obtained extensions of his earlier results.

Finally I would like to mention the deep and difficult work of E. Heinz in a series of papers [18–20] on elliptic Monge–Ampère equations in the plane

$$u_{xx} u_{yy} - u_{xy}^2 + \text{quasilinear second order expressions} = 0.$$

In these papers Heinz extends the important work of Hans Lewy on such equations which are analytic to non analytic equations, deriving estimates for derivatives of solutions. This work really deserves more time here but the results are complicated to state. Let me just say that Heinz makes essential use of characteristic coordinates for the equation, i.e. new coordinates (depending on the solution) which reduce the leading part of the equation regarded as linear to the Laplace operator.

To mention just one of his simpler results, proved with the aid of characteristic coordinates and some theorems on quasi-conformal mappings, he shows [21] that if

$$0 < \alpha \leqslant u_{xx} u_{yy} - u_{xy}^2 \leqslant \beta < \infty \quad \text{and} \quad |u| \leqslant \delta$$

in a circle then in any compact subset the first derivatives of u (which are easily estimated a priori because of convexity of $u = u(x, y)$) satisfy also a fixed Hölder condition.

I might also mention that in [17] Heinz proved the existence of surfaces of constant mean curvature spanning a given closed curve in space.

§ 5. I have said nothing about nonlinear hyperbolic equations. Recently using extensions of Sobolev's inequalities L. Gårding and Leray [15] and P. A. Dionne [7] have treated the existence and uniqueness of regular solu-

tion for the initial value problem and have reduced it to the possibility of deriving a priori estimates of a certain number of derivatives of the solutions —in case one wants global solutions.

Besides that much work has been done in recent years on the very important initial value problem involving discontinuous initial data. Here, in contrast to elliptic problems one seeks some form of generalized solution admitting discontinuities—shocks to be exact. The basic problem of finding the appropriate class of generalized solution in which one has existence and uniqueness has still not been solved except in special, though very interesting cases. In particular in the case of more than one space variable essentially nothing is known. There are expository papers on the subject by P. D. Lax [31], O. A. Oleinik [49], and B. L. Rozhdestvenskii [53].

REFERENCES

[1]. Agmon, S., Douglis, A. & Nirenberg, L., Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I (II to appear). *Comm. Pure Appl. Math.*, 12 (1959), 623–727.

[2]. Bers, L. & Nirenberg, L., On a representation theorem for linear elliptic systems with discontinuous coefficients and its application. On linear and nonlinear elliptic boundary value problems in the plane. *Convegno Int. Equazione Lin. alle Derivate Parziale.* Trieste, August 1954. Cremonese, Rome (1955), 111–140, 141–167.

[3]. Browder, F., *A priori* estimates for solutions of elliptic boundary value problems I, II. *Nederl. Akad. Wetensch. Indag. Math.* 22 (1960), 145–159, 160–169.

[4]. Calderón, A. P., Intermediate spaces and interpolation. *Warsaw Meeting*, summer 1960.

[5]. Cordes, H. O., Zero order a priori estimates for solutions of elliptic differential equations. Proc. Symposium in Pure Math. A.M.S. Partial dif. ferential Equations Vol. 4, held at Berkely, California, April 1960-*Amer. Math. Soc.* (1961), 157–166.

[6]. de Giorgi, E., Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari. *Mem. Accad. Sci. Torino* ser. 3, part 1 (1957), 25–43.

[7]. Dionne, P. A., Problème de Cauchy hyperbolique et espaces fonctionnels de S. L. Sobolev. *Coll. Int. CNRS sur les Equations aux Dérivées Partielles.* Paris, June 1962.

[8]. Filippov, A. F., On conditions of existence of the solution for a quasilinear parabolic equation. *Dokl. Akad. Nauk, SSSR* 141 (1961), 568–570.

[9]. Finn, R., On equations of minimal surface type. *Ann. Math.* 60 (1954), 397–416.

[10]. Finn, R., Estimates at infinity for steady state solutions of the Navier-Stokes equations. *Amer. Math. Soc. Symp.*, Vol. 4 on Partial Differential Equations, Berkeley (April 1960), 143–148.

[11]. Friedman, A., Interior estimates for parabolic systems of partial differential equations. Boundary estimates for second order parabolic equations and their applications. *J. Math. Mech.*, 7 (1958), 393–418; 771–792.

[12]. —— On quasilinear parabolic equations of the second order I and II. *J. Math. Mech.*, 7 (1958), 793–809 and 9 (1960), 539–556.

[13]. GAGLIARDO, E., Ulteriori proprietà di alcune classi di funzioni in piu variabili. *Ricerche Mat.*, 8 (1959), 24–51.

[14]. —— Interpolazione di spazi di Banach e applicazioni, *Ricerche Mat.*, 9 (1960), 58–81.

[15]. GÅRDING, L., Problème de Cauchy pour les systèmes quasi-linéaires d'ordre un strictement hyperboliques. *Coll. Int. du CNRS sur les Equations aux Dérivées Partielles*, Paris, June 1962.

[16]. GILBARG, D., Boundary value problems for nonlinear elliptic equations in n variables. *Symp. on Non-Linear Problems. Math. Research Center, U.S. Army.* University of Wisconsin, Madison, Wisconsin. Spring 1962.

[17]. HEINZ, E., Über die Existenz einer Fläche konstanter mittlerer Krümmung bei vorgegebener Berandung. *Math. Ann.*, 127 (1954), 258–287.

[18]. —— On certain nonlinear elliptic differential equations and univalent mappings. *J. Analyse Math.*, 5 (1956–7), 197–272.

[19]. —— On elliptic Monge-Ampère equations and Weyl's embedding problem. *J. Analyse Math.*, 7 (1959), 1–52.

[20]. —— Neue a priori Abschätzungen für den Ortsvektor einer Fläche positiver Gauss'scher Krümmung durch ihr Linienelement. *Math. Z.*, 74 (1960), 129–157.

[21]. —— Über die Differentialungleichung $0 < \alpha \leqslant rt - s^2 \leqslant \beta < \infty$. *Math. Z.*, 72 (1959), 107–126.

[22]. HÖRMANDER, L., On the regularity of the solutions of boundary problems. *Acta Math.*, 99 (1958), 225–264.

[23]. JENKINS, H. B., On two-dimensional variational problems in parametric form. *Arch. Rational Mech. Anal.* 8, no. 3 (1961), 181–206.

[24]. JOHN, F., Rotation and strain. *Comm. Pure Appl. Math.*, 14 (1961), 391–414.

[25]. JOHN, F. & NIRENBERG, L., On functions of bounded mean oscillation. *Comm. Pure Appl. Math.*, 14 (1961), 415–426.

[26]. KAPLAN, S., On the growth of solutions of quasilinear parabolic equations. (To appear in *Comm. Pure Appl. Math.*)

[27]. KOHN, J. J., Solution of the $\bar{\partial}$-Neumann problem on strongly pseudoconvex manifolds. *Proc. Nat. Acad. Sci.*, 47 (1961). 1198–1202.

[28]. LADYZHENSKAYA, O. A., Mathematical problems of viscous non-compressible fluid flow. *Moscow Gos. Izd. Fiz. Mat.* Lit. 1961.

[29]. LADYZHENSKAYA, O. A. & URALTSEVA, N. N., Quasilinear elliptic equations and variational problems with many independent variables. *Uspehi Mat. Nauk* 16. no. 1 (1961), 19–92 transl. *Russian Math. Surveys, London Math. Soc.*, 16 (1) (1961) 17–91.

[30]. LADYZHENSKAYA, O. A. & URALTSEVA, N. N., Boundary value problems for linear and quasilinear parabolic equations I. *Izv. Akad. Nauk SSSR*, 26 (1), (1962), 5–52.

[31]. LAX, P. D., Nonlinear hyperbolic systems of conservation laws. *Symp. on Non-Linear Problems. Math. Research Center, U.S. Army.* University of Wisconsin, Madison, Wisconsin. Spring 1962.

[32]. LERAY, J., La théorie des points fixes et ses applications en analyse, *Proc. Internat. Congress Math. Cambridge, Mass.* II (1950), 202–208.

[33]. LERAY, J. & SCHAUDER, J., Topologie et équations fonctionelles. *Ann. Sci. Ec. Norm. Sup.* Sér. 3, 51 (1934), 45–78.

[34]. LIONS, J. L., Sur les espaces d'interpolation, dualité. *Math. Scand.*, 9 (1961), 147–177.

[35]. —— *Equations Differentielles Operationnelles.* Springer, 1961.

[36]. MIRANDA, C., *Equazioni alle derivate parziali di tipo ellittico.* Springer, 1955.

[37]. MORREY, C. B., On the solutions of quasilinear elliptic partial differential equations. *Trans. Amer. Math. Soc.*, 43 (1938), 126–166.

[38]. MORREY, C. B., Multiple integral problems in the calculus of variations and related topics, *Univ. California Publ. Math.*, N.S., 1 (1943), 1–130.

[39]. —— The analytic embedding of abstract real analytic manifolds. *Ann. Math.*, 68 (1958), 159–201.

[40]. —— *Existence and differentiability theorems for variational problems for multiple integrals. Partial differential equations and continuum mechanics.* Univ. of Wisconsin Press, Madison, Wisc., 1961, 241–270.

[41]. —— Quelques résultats récents du calcul des variations. *Coll. Int. du CNRS sur les Equations aux Dérivées Partielles.* Paris, June 1962.

[42]. MOSER, J., A new proof of de Giorgi's theorems concerning the regularity problem for elliptic differential equations. *Comm. Pure Appl. Math.*, 13 (1960), 457–468.

[43]. —— On Harnack's theorem for elliptic differential equations. *Comm. Pure Appl. Math.*, 14 (1961), 577–591.

[44]. —— A new method for the construction of solutions for non-linear problems in analysis. *Proc. Nat. Acad. Sci.*, 47 (11) (1961), 1824–1831.

[45]. NAGUMO, M., Degree of mapping in convex linear topological spaces. *Amer. J. Math.*, 73 (1951), 497–511.

[46]. NASH, J., The imbedding problem for Riemannian manifolds. *Ann. Math.*, 63 (1956), 20–63.

[47]. —— Continuity of the solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80 (1958), 931–954.

[48]. NIRENBERG, L., On elliptic partial differential equations. *Ann. Scuola Norm. Sup. Pisa*, Ser 3, 13 (1959), 116–162.

[49]. OLEINIK, O. A., Discontinuous solutions of non-linear differential equations. *Uspehi Mat. Nauk*, 12 (1957), 3–73.

[50]. OLEINIK, O. A. & KRUZHKOV, C. N., Quasilinear parabolic equations of second order with many independent variables. *Uspehi Mat. Nauk*, 16 (5), (1961), 115–155.

[51]. OSSERMAN, R., On the Gauss curvature of minimal surfaces. *Trans. Amer. Math. Soc.*, 96 (1960), 115–128.

[52]. PRODI, G., Résultats récents et problèmes anciens dans la theorie des équations de Navier-Stokes. *Coll. Int. du CNRS sur les Équations aux Dérivées Partielles.* Paris, June 1962.

[53]. ROZHDESTVENSKII, B. L., Discontinuous solutions of hyperbolic systems of quasilinear equations. *Uspehi Mat. Nauk*, 15 (1959) 53–111., trans. *Russian Math. Surveys, London Math. Soc.*, 15 (5) (1960) 53–111.

[54]. SCHWARTZ, J., On Nash's implicit functional theorem. *Comm. Pure Appl. Math.*, 13 (1960), 509–530.

[55]. SERRIN, J., The initial value problem of the Navier-Stokes equations. *Symp. on Non-Linear Problems, Math. Research Center, U.S. Army.* University of Wisconsin, Madison, Wisconsin. Spring 1962.

[56]. STAMPACCHIA, G., Problemi al contorno ellittici con dati discontinuiti

dotati di soluzioni hölderiane. *Ann. Mat. Pura Appl.*, (IV) 51 (1960), 1–38.

[57]. —— On some regular multiple integral problems in the calculus of variations. (To appear in *Comm. Pure Appl. Math.*)

[58]. Temple, G., Linearization and delinearization. *Proc. Internat. Congress Math. Edinburgh*, 1958, 233–247.

[59]. Vishik, M. I., Boundary value problems for quasilinear strongly elliptic systems of divergence form. *Dokl. Akad. Nauk. SSSR*, 138 (1961) 518. *Translation Soviet Math. Dokl.*, 2 (1961), 643–647.

[60]. —— Systems of quasilinear elliptic equations with lower order terms. *Dokl. Akad. Nauk. SSSR*, 144 (1962). *Translation Soviet Math. Dokl.*, 3 (1962), 629–633.

ПОЛЯ АЛГЕБРАИЧЕСКИХ ЧИСЕЛ

И. Р. ШАФАРЕВИЧ

Введение

Теория алгебраических чисел имеет две стороны. Одна из них — это построение общей теории полей алгебраических чисел, имеющее целью их полную классификацию и описание их арифметики. Другая сторона связана с приложением этой общей теории к таким конкретным вопросам, как, например, диофантовы уравнения, комплексное умножение эллиптических и абелевых функций или интегрирование алгебраических дифференциалов.

В этом обзоре будут рассмотрены некоторые вопросы, лежащие в первом направлении.

Образцом для всех будущих теорий здесь служит теория полей классов, которая исчерпывающим образом описывает все абелевы (имеющие коммутативную группу Галуа) расширения фиксированного поля алгебраических чисел. В области расширений с неабелевой группой Галуа наши знания очень далеки от такой полноты. Часто неясно даже, как следует ставить вопросы и в каких терминах искать их решение. Иногда легче найти правильную постановку вопроса, если посмотреть на эту область с более общей точки зрения.

Поля алгебраических чисел и содержащиеся в них кольца целых алгебраических чисел являются простейшими примерами и в то же время отправными пунктами для гораздо более общих теорий. Особенно большое значение эти связи имеют для самой теории алгебраических чисел, так как они помогают найти новые постановки вопросов, сопоставляя эту теорию с другими, которые ей во многом аналогичны, но в которых мы можем использовать аналитический аппарат или геометрическую интуицию. Это объясняет, почему дальше речь будет идти часто о вопросах, не относящихся, строго говоря, к теории алгебраических чисел.

§ 1. Поля алгебраических функций

Особенно большую роль в развитии теории алгебраических чисел играли ее связи и аналогии с теорией алгебраических функций одной переменной. Пусть k — поле алгебраических функций одной переменной с полем констант k_0. Нас интересует проблема классификации конечных расширений поля k. Решение ее хорошо известно, если $k_0=C$ — поле комплексных чисел. В этом случае полю k может быть сопоставлена его риманова поверхность $\mathfrak{R}$. Если k' — конечное расширение k, то его риманова поверхность $\mathfrak{R}'$ естественным образом непрерывно отображается на $\mathfrak{R}$: $\mathfrak{R}'\to\mathfrak{R}$. Таким образом, поверхность $\mathfrak{R}^1$ является накрывающей для поверхности $\mathfrak{R}$. Это накрытие имеет ту же степень, что и расширение k'/k и обладает конечным числом точек ветвления. Из тео-

ремы существования Римана легко вытекает, что мы получаем таким образом взаимно однозначное соответствие между конечными расширениями поля k и разветвленными накрытиями конечной степени его римановой поверхности $\mathfrak{R}$. Задача сводится, таким образом, к топологическому вопросу, ответ на который хорошо известен. Фиксируем конечное множество $S=\{\mathfrak{p}_1, \ldots, \mathfrak{p}_m\}$ точек поверхности $\mathfrak{R}$ и рассмотрим накрытия конечной степени $\mathfrak{R}' \to \mathfrak{R}$, разветвленные только в точках этого множества. Они соответствуют неразветвленным накрытиям поверхности $\mathfrak{R}-S$, которые, в свою очередь, находятся во взаимно однозначном соответствии с подгруппами конечного индекса фундаментальной группы $\pi_1(\mathfrak{R}-S)$ этой поверхности. Наконец, группа $\pi_1(\mathfrak{R}-S)$ хорошо известна — она имеет $m+2g$ образующих $u_1 \ldots, u_m, s_1, t_1, \ldots s_g, t_g$ (g — род поверхности $\mathfrak{R}$), связанных одним соотношением

$$u_1 \ldots u_m s_1 t_1 s_1^{-1} t_1^{-1} \ldots s_g t_g s_g^{-1} t_g^{-1} = 1. \tag{1}$$

Заметим, что хотя источники этих результатов аналитические, им может быть придана алгебраическая формулировка. Действительно, точки поверхности $\mathfrak{R}$ соответствуют простым дивизорам поля k и, в частности, точки ветвления накрытия $\mathfrak{R}' \to \mathfrak{R}$ могут быть характеризованы в терминах расширения k'/k как критические простые дивизоры (делители дискриминанта). Группа $\pi_1(\mathfrak{R}-S)$ также может быть описана в алгебраических терминах. Для этого рассмотрим поле $\mathfrak{K}_S$, являющееся объединением всех расширений k'/k, критические простые дивизоры которых принадлежат множеству S. Это расширение бесконечно, и его группа Галуа $\mathfrak{G}_S$ является топологической группой. Таким образом, группа $\mathfrak{G}_S$ может быть определена для произвольного поля алгебраических функций. Ее знание дает возможность перечислить все расширения k'/k с указанным выше свойством, так как они, согласно основной теореме теории Галуа, соответствуют ее подгруппам конечного индекса. Если $k_0=C$, то $\mathfrak{G}_S$ является пополнением группы $\pi_1(\mathfrak{R}—S)$ в топологии, определенной всеми подгруппами конечного индекса. Таким образом, соотношение (1) определяет $\mathfrak{G}_S$ как топологическую группу.

Известно, что среди полей алгебраических функций к полям алгебраических чисел ближе всего стоят те, у которых поле k_0 конечно. Ввиду этого с точки зрения теории полей алгебраических чисел было бы интересно получить для этого класса полей результат, аналогичный сформулированному выше. Решение, однако, известно только в промежуточном случае, когда поле констант k_0 алгебраически замкнуто.

Если характеристика поля k_0 равна 0, то можно применить известный «принцип Лефшеца». Задача легко сводится к случаю, когда поле коэффициентов k_0 имеет конечную степень трансцендентности над полем рациональных чисел, а путем вложения такого поля в поле комплексных чисел — к разобранному случаю $k_0=C$. Из этих соображений следует, что для любого поля алгебраических функций с алгебраически замкнутым полем констант k_0 характеристики 0 группа $\mathfrak{G}_S$ такая же, как и в случае $k_0=C$.

Для алгебраически замкнутого поля k_0 характеристики $p>0$ аналогичный результат получен А. Гротендиком [1] для расширений, степень которых взаимно проста с p. Для описания этого результата построим поле $\mathfrak{L}_S$, так же как раньше строилось поле $\mathfrak{K}_S$ ограничиваясь, однако,

теперь только нормальными полями, степени которых взаимно просты с p. Оказывается, что группа Галуа $\mathfrak{H}_S$ поля $\mathfrak{L}_S$ опять та же, что и в классическом случае $k_0=C$. Доказательство основывается на сведении к случаю поля характеристики 0 путем редукции таких полей по модулю p. Таким образом в основе доказательства лежит такой глубокий аналитический факт как теорема существования Римана.

Вернемся теперь к случаю конечного поля констант k_0, который для нас особенно интересен. Обозначим через $\bar{k}_0$ алгебраическое замыкание поля k_0 и через $\bar{k}$ — композит полей k и $\bar{k}_0$. Очевидно, что $k\subset\bar{k}\subset\mathfrak{L}_S$. Группа Галуа $\mathfrak{H}_S$ поля $\mathfrak{L}_S/\bar{k}$ описана теоремой Гротендика, а группа Галуа $\mathfrak{F}$ поля $\bar{k}/k$ хорошо известна — она является свободной топологической группой с одной образующей. Так как $\mathfrak{H}_S/\overline{\mathfrak{H}}_S=\mathfrak{F}$, то определение группы $\mathfrak{H}_S$ — это задача теории расширений групп. Ввиду того, что группа $\mathfrak{F}$ свободная, соответствующее расширение распадается и, следовательно, структура группы $\mathfrak{H}_S$ определяется тем, как образующий элемент группы $\mathfrak{F}$ (так называемый автоморфизм Фробениуса) действует на группу $\mathfrak{H}_S$. К сожалению, об этом ничего не известно, даже в самых простых случаях (например, в случае p-расширений).

§ 2. Локальные поля

Прежде чем приступить к исследованию полей алгебраических чисел, естественно исследовать локальные поля, т. е. пополнения полей алгебраических чисел по их простым дивизорам.

Пусть k — локальное поле, содержащее поле p-адических чисел R_p. Согласно теории ветвления Дедекинда–Гильберта в любом нормальном расширении K/k содержится максимальное подполе k_1/k, не имеющее высшего ветвления. Степень поля K/k_1 является степенью p. Таким образом, всякое расширение локального поля в некотором смысле составлено из поля без высшего ветвления и p-расширения. Структура полей без высшего ветвления весьма проста, как следует из той же теории Дедекинда–Гильберта. Поэтому можно думать, что основная трудность в классификации локальных полей — это классификация p-расширений.

Сейчас имеется довольно полная картина строения p-расширений локального поля. Для их описания рассмотрим аналогично тому, как мы это делали в случае полей алгебраических функций, поле $\mathfrak{K}$, являющееся объединением всех конечных p-расширений локального поля k и обозначим через $\mathfrak{G}$ группу Галуа расширения $\mathfrak{K}/k$. Группа $\mathfrak{G}$ является топологической p-группой и содержит в себе довольно полную информацию о p-расширениях поля k, так как они находятся во взаимно однозначном соответствии с подгруппами конечного индекса группы $\mathfrak{G}$. Поэтому основные усилия были направлены на вычисление этой группы.

Обозначим через ν степень $[k:R_p]$. Строение группы $\mathfrak{G}$ зависит от этого числа и от того, содержится ли в поле k первообразный корень ζ степени p из единицы.

Имеют место следующие результаты:

1. Если $\zeta\notin k$, то группа $\mathfrak{G}$ является свободной топологической p-группой с $\nu+1$ образующими [2].

2. Если $\zeta\in k$, то группа $\mathfrak{G}$ имеет $\nu+2$ образующие и определена (как

топологическая p-группа) одним определяющим соотношением ([3] и [4]).

В этих результатах утверждения о числе образующих легко следуют из известной теоремы Бернсайда [5].

Доказательство утверждений о числе соотношений основывается на соображениях, которые мы рассмотрим сразу в более общей ситуации.

Пусть k — произвольное поле и $\mathfrak{K}$ его p-расширение (конечное или нет) с группой Галуа $\mathfrak{G}$. Легко видеть, что число r независимых соотношений, связывающих минимальную систему образующих группы $\mathfrak{G}$ равно числу образующих ее двумерной группы гомологий:

$$r = \dim_{\mathbf{Z}_p} H^2(\mathfrak{G}, \mathbf{Z}_p),$$

где $\mathbf{Z}_p$ группа порядка p.

Пусть K_α — семейство подполей $\mathfrak{K}$, конечных над k и G_α — их группы Галуа. Определены гомоморфизмы

$$j_\alpha: H^2(G_\alpha, \mathbf{Z}_p) \to H^2(\mathfrak{G}, \mathbf{Z}_p)$$

и $H^2(\mathfrak{G}, \mathbf{Z}_p)$ является объединением групп $j_\alpha H^2(G_\alpha, \mathbf{Z}_p)$. Каждый элемент $\xi \in H^2(G_\alpha, \mathbf{Z}_p)$ определяет некоторое расширение G_α^ξ группы G_α при помощи $\mathbf{Z}_p$. Легко видеть, что $j_\alpha \xi = 0$ тогда, и только тогда, когда поле K_α погружаемо в поле $K' \subset \mathfrak{K}$, имеющее группу Галуа G_α^ξ. Это дает возможность определить группу $H^2(\mathfrak{G}, \mathbf{Z}_p)$ (а, значит, и число соотношений r в группе $\mathfrak{G}$), если мы умеем решать соответствующую задачу погружения. Несколько неточно можно сказать, что r равно числу условий, которые должны овыплняться, чтобы эта задача погружения была разрешима. В частности, результат 2 непосредственно вытекает из этих соображений, так как условием разрешимости задачи погружения является, согласно известному критерию Брауера [6] обращение в 1 некоторого элемента группы $B(k)_p$, где $B(k)$ — группа Брауера поля k, а $B(k)_p$ — ее подгруппа, состоящая из элементов периода p. Известно, что для любого n группы $B(k)_n$ имеет одну образующую. Отсюда и следует, что $r = 1$.

Результат 1 описывает группу $\mathfrak{G}$ полностью. В случае 2 остается найти соотношение, определяющее группу $\mathfrak{G}$. Это было сделано С. П. Демушкиным [7]. Оказалось, что если μ — наибольшее число, такое что поле k содержит первообразный корень степени p^μ из 1, то соотношение имеет вид (при $p \neq 2$ или $p = 2$, $\mu \neq 1$)

$$S_1^{p^\mu} (s_1, s_2)(s_3, s_4) \ldots (s_{\nu+1}, s_{\nu+2}) = 1, \tag{2}$$

где (s, t) означает $s\,t\,s^{-1}\,t^{-1}$. Случай $p = 2$, $\mu = 1$ (который имеет место, например, если $k = R_2$) до сих пор не разобран.

Этот результат носит, по существу, теоретико-групповой характер и связан с интересным вопросом — исследованием топологических p-групп, определенных одним соотношением. Соотношение f, определяющее некоторую p-группу G, называется полным, если его нельзя преобразовать подходящей заменой образующих в такое соотношение, которое не содержит какой-либо из образующих. Это требование равно-

сильно тому, что группа G не разлагается в свободное произведение. Соотношение f называется m-полным, если то же имеет место для группы $G/G^{(m)}$ где $G^{(m)}$ — m-й член убывающего центрального ряда группы G.

Для группы $\mathfrak{G}$ (группы Галуа максимального p-расширения локального поля k) нетрудно вычислить группу $\mathfrak{G}/\mathfrak{G}^{(2)}$ и проверить, что соотношение, определяющее $\mathfrak{G}$ 2-полно. Демушкин доказывает для произвольной p-группы G, определенной одним соотношением f (ограничиваясь, правда, группами с четным числом образующих), что если f 2-полно, то группа G однозначно определяется группой $G/G^{(2)}$. Отсюда легко вытекает классификация всех соотношений такого типа — они всегда могут быть приведены к виду (2) с некоторым μ. То, что для группы $\mathfrak{G}$ число μ имеет указанный выше смысл, получается из рассмотрения группы $G/G^{(1)}$.

Было бы очень интересно продолжать эти исследования и выяснить, как обобщаются эти результаты при переходе к m-полным соотношениям при произвольном m. Отметим одно теоретико-групповое следствие этой теоремы о локальных полях — подгруппа конечного индекса p-группы, определенной соотношением (2), сама определена одним соотношением такого же типа (возможно, с другим μ и ν). Естественно возникает вопрос, обобщается ли этот факт на любой p-группы с одним соотношением и можно ли в этом случае придать какой-нибудь смысл словам «соотношение того же типа».

Естественно надеяться, что изучение любых конечных расширений локальных полей может быть сведено к изучению расширений без высшего ветвления и p-расширений. Этот вопрос равносилен вычислению группы Галуа алгебраического замыкания локального поля. Задача еще не решена, хотя Х. Кох (см. [8] и диссертацию, представленную в Университет им. Гумбольдта в Берлине) нашел ряд свойств этой группы и доказал, что она однозначно определяется простыми инвариантами поля k.

§ 3. Поля алгебраических чисел

Естественно попытаться использовать в теории полей алгебраических чисел тот же подход, которым мы пользовались в полях алгебраических функций и в локальных полях. Пусть k — некоторое поле алгебраических чисел, конечное расширение поля рациональных чисел R. Обозначим через S некоторое конечное множество простых дивизоров поля k и рассмотрим все конечные разширения k'/k, критическими простыми дивизорами (т. е. делителями дискриминантов) которых являются только простые дивизоры множества S. Объединение всех этих полей обозначим через $\mathfrak{K}_S$, а группу Галуа поля $\mathfrak{K}_S/k$ через $\mathfrak{G}_S$. Повидимому, одной из основных задач теории алгебраических чисел является исследование групп $\mathfrak{G}_S$.

По аналогии с тем, что, как мы видели, имеет место в полях алгебраических функций, возникают два вопроса:

Будет ли группа $\mathfrak{G}_S$ всегда обладать конечной системой образующих?

Если это так, то можно ли (при фиксированном поле k) ограничить число образующих в зависимости только от числа дивизоров в множестве S?

Ответы на эти вопросы не известны. Положительный ответ на второй из них означал бы, что для любого конечного нормального расширения K/k число образующих его группы Галуа ограничено в зависимости от числа критических простых дивизоров. Этот вопрос интересен уже для самых простых расширений. Пусть, например, k совпадает с полем рациональных чисел R. Обозначим через L/R квадратичное расширение с простым дискриминантом p и возьмем за K Гильбертово поле классов поля L (т. е. его максимальное абелево неразветвленное расширение). Очевидно, что K/R имеет только одно критическое простое число p. Так как группа Галуа K/L совпадает с группой классов дивизоров L, а число образующих группы Галуа K/R отличается, как легко видеть, от числа образующих группы Галуа K/L не больше, чем на 1, то интересующий нас вопрос эквивалентен в этом случае следующему:

Ограничены ли числа образующих групп классов дивизоров квадратичных полей с простым дискриминантом?

Этим кругом вопросов интересовался еще Гаусс, который называл квадратичное поле регулярным, если его группа классов дивизоров циклическая [9].

По своему обыкновению Гаусс составил таблицу [9], из которых следует, что для $p<3000$ и для $9000<p<10\,000$ число образующих равно 1 или 2, причем во втором случае одна образующая имеет порядок 3.

Уже этот простой пример показывает, что исследование групп $\mathfrak{G}_S$ связано с очень большими трудностями. Повидимому, в настоящее время в этом направлении не известно никаких общих результатов. Можно сказать гораздо больше, если ограничиться рассмотрением p-расширений. Фиксируем некоторое простое число p и обозначим через $\mathfrak{P}_S$ группу Галуа объединения всех p-расширений поля k, критические простые дивизоры которых содержатся в множестве S. Очевидно, что группа $\mathfrak{P}_S$ является максимальной p-факторгруппой группы $\mathfrak{G}_S$.

Легко видеть, что оба вопроса, поставленные выше для групп $\mathfrak{G}_S$, имеют для групп $\mathfrak{P}_S$ положительный ответ.

Обозначим минимальное число образующих группы $\mathfrak{P}_S$ через $d(S)$. Это число является довольно тонким инвариантом поля k и простых дивизоров, составляющих множество S.

По порядку величины $d(S)$ близко числу элементов множества S. Точнее говоря, существуют такие две постоянные c_1 и c_2, зависящие только от поля k, что

$$c_1 \leqslant d(S) - m \leqslant c_2,$$

где m обозначает число простых дивизоров, входящих в S для которых $\zeta \in k_y$.

Особенно интересен случай, когда множество S пусто. Группу $\mathfrak{P}_S$ мы будем тогда обозначать просто через $\mathfrak{P}$. Она является, очевидно, группой Галуа максимального неразветвленного p-расширения поля k. Из теории полей классов следует, что все факторы ряда коммутантов группы $\mathfrak{P}$ конечны. Обрывается ли этот ряд, т. е. конечна ли группа $\mathfrak{P}$ — неизвестно. Этот вопрос составляет содержание так называемой проблемы p-башни.

Близкий случай, когда все простые дивизоры из S взаимно просты с p. Здесь тоже факторы ряда коммутантов конечны. Вопрос об обрыве ряда близок к тому же вопросу для группы $\mathfrak{P}$.

Аналогично тому, как мы поступали в теории полей алгебраических функций и в теории локальных полей, естественно попытаться найти соотношения, определяющие группы $\mathfrak{P}_S$. Удается оценить сверху их число. Это можно сделать пользуясь тем общим методом определения числа соотношений в p-группе, который был описан в § 2. Определение числа соотношений связано, как мы видели, с решением некоторой задачи погружения. В нашем случае это задача погружения поля, все критические простые дивизоры содержатся в S, в поле такого же типа. Исследуя эту задачу, можно получить следующий результат:

Число соотношений $r(\mathfrak{P}_S)$, определяющих группу $\mathfrak{P}_S$ удовлетворяет неравенству

$$r(\mathfrak{P}_S) \leqslant \gamma + m + c,$$

где γ обозначает число образующих p-компоненты группы классов поля k, m — число простых дивизоров в S, для которых $\zeta \in k_y$, а c — постоянная, зависящая только от степени $[k : R]$.

Например, для группы $\mathfrak{P}$ мы получаем неравенство

$$r(\mathfrak{P}) \leqslant \gamma + c.$$

Легко видеть, что $d(\mathfrak{P}) = \gamma$ так, что

$$r(\mathfrak{P}) \leqslant d(\mathfrak{P}) + c. \tag{3}$$

Из грубых теоретико-групповых соображений следует, что если для некоторого семейства p-групп G, $r(G) \leqslant d(G) + c$, где c — постоянная, то класс групп G стремится к бесконечности вместе с $d(G)$. Отсюда вытекает следующий результат:

Для полей k фиксированной степени класс группы $\mathfrak{P}$ неограниченно возрастает вместе с числом образующих p-компоненты группы классов поля k.

Шольц [10] построил некоторые примеры полей k, у которых группы $\mathfrak{P}$ имеют сколь угодно большой класс. Из предшествующего результата следует, что это на самом деле общая закономерность. Гипотезу (несколько более частного характера) высказывал Фрелих [11].

Возникает, однако, чувство, что число соотношений, удовлетворяющее неравенствы (3), при $d(\mathfrak{P})$ достаточно большом, сравнительно с c, вообще слишком мало, чтобы определить конечную p-группу. Действительно, самая простая группа — прямое произведение d групп $\mathbf{Z}_p$ определяется $d(d+1)/2$ соотношениями

$$\sigma_i^p = 1, i = 1, \ldots, d; \sigma_i \sigma_j = \sigma_j \sigma_i \quad (1 \leqslant i < j \leqslant d).$$

В связи с этим мне кажется очень принципиальным следующий вопрос, к которому я хотел бы привлечь внимание специалистов по теории групп:

Каково наименьшее число соотношений, необходимое для того, чтобы определить конечную p-группу с d образующими?

Поставим этот вопрос точно. Определим функцию $\varrho(d)$ следующим образом:

$$\varrho(d)=\min_{d(G)=d} r(G),$$

где минимум берется по всем конечным группам, минимальное число образующих которых равно d. Речь идет об оценке снизу функции $\varrho(d)$ при росте d. Из рассмотрения факторгруппы по коммутанту следует, что $\varrho(d)\geqslant d$. Если бы можно было показать, что при $d\to\infty$

$$\varrho(d)-d\to\infty,$$

что мне кажется правдоподобным, то отсюда следовало бы отрицательное решение проблемы p-башни со всеми вытекающими из этого следствиями: существованием полей, не вкладывающихся в поля с однозначным разложением на множители, оценкой истинного порядка роста минимального дискриминанта полей алгебраических чисел и т. д. Так как $r(G)=\dim_{\mathbf{Z}_p} H^2(G,\mathbf{Z}_p)$, $d(G)=\dim_{\mathbf{Z}_p} H^1(G,\mathbf{Z}_p)$, то этот вопрос эквивалентен вопросу об оценке порядка второй группы гомологий конечной p-группы при фиксированной первой группе гомологий.

В заключение заметим, что все сказанное относится также и к группам $\mathfrak{P}_S$, если все простые дивизоры множества S взаимно просты с p.

§ 4. Критические простые дивизоры алгебраических многообразий над полями алгебраических чисел

Вопросы, связанные с критическими простыми дивизорами полей алгебраических чисел, могут быть поставлены и для алгебраических многообразий, определенных над полями алгебраических чисел и приводят нас к интересным задачам и к некоторым результатам.

До сих пор мы не формулировали определение критического простого дивизора расширения k'/k. Чтобы привести это определение, рассмотрим простой дивизор $\mathfrak{p}$ поля k, его локальное кольцо $\sigma_{\mathfrak{p}}$ и целое замыкание $\sigma'_{\mathfrak{p}}$ кольца $\sigma_{\mathfrak{p}}$ в k'. Кольцо $\sigma'_{\mathfrak{p}}/\mathfrak{p}$ является алгеброй над полем $k(\mathfrak{p})=\sigma_{\mathfrak{p}}/\mathfrak{p}$. Если эта алгебра сепарабельна, то $\mathfrak{p}$ называется некритическим простым дивизором, в противном случае — критическим.

Рассмотрим теперь полное неособое алгебраическое многообразие X, определенное над полем алгебраических чисел k и простой дивизор $\mathfrak{p}$ поля k. Редукция $X/\mathfrak{p}$ многообразия X по модулю $\mathfrak{p}$ (определение которой дано, например, в [12]) является алгебраическим множеством над полем $k(\mathfrak{p})$. $X/\mathfrak{p}$ может уже иметь особые точки. Это значит, что локальные кольца некоторых его точек не будут регулярными. В случае кольца $\sigma'_{\mathfrak{p}}/\mathfrak{p}$ условие регулярности локальных колец его простых идеалов совпадает с сепарабельностью. Это приводит нас к следующему определению: простой дивизор $\mathfrak{p}$ поля k называется некритическим для полного неособого многообразия X, определенного над k, если $X/\mathfrak{p}$ — неособое многообразие над полем $k(\mathfrak{p})$; в противном случае $\mathfrak{p}$ называется критическим. В [12] доказано, что X имеет только конечное число критических простых дивизоров.

Понятию критического простого дивизора удобно придать бирационально инвариантную форму, связав его с полем алгебраических функций, а не с многообразием X-моделью этого поля. Пусть K — поле

рациональных функций на некотором полном неособом алгебраическом многообразии X, определенном над k. Простой дивизор $\mathfrak{p}$ поля k называется некритическим для поля K, если существует хотя бы одно неособое полное многообразие X', для которого K является полем функций, а $\mathfrak{p}$ — некритическим простым дивизором в смысле предшествующего определения.

Одна из основных теорем алгебраических чисел, связанная с понятием дискриминанта — это теорема Эрмита, утверждающая, что число расширений k'/k с заданной степенью и с заданным дискриминантом конечно. Этой теореме можно придать следующую форму: число расширений k'/k заданной степени, критические простые дивизоры которых принадлежат заданному конечному множеству S, конечно.

В такой форме это утверждение может быть сформулировано, хотя бы как вопрос, для алгебраических многообразий любой размерности n или соответствующих полей алгебраических функций. Теорема Эрмита соответствует случаю $n=0$. Мы рассмотрим следующий по сложности случай $n=1$, т. е. случай алгебраических кривых или полей алгебраических функций K одной переменной. Мы будем считать, что поле K определено над полем алгебраических чисел k и что k алгебраически замкнуто в K. При этих предположениях имеет место следующий результат, аналогичный теореме Эрмита:

Существует только конечное число полей алгебраических функций K/k заданного рода $g \neq 1$, критические простые дивизоры которых принадлежат заданному конечному множеству S.

Этот результат имеет место, если его надлежащим образом модифицировать, и для полей рода $g=1$. Для этого нужно, кроме рода, рассмотреть другой важный инвариант полей алгебраических функций рода 1 — общий наибольший делитель d степеней всех дивизоров поля, называемый показателем поля. Известно [13], что показатель может принимать сколь угодно большие значения. Если ограничиться рассмотрением полей рода 1 с заданным показателем d, то сформулированная выше теорема конечности остается верной. Например, при $k=R$ и в простейшем случае $d=1$ (когда K является полем функций на эллиптической кривой) имеется 24 поля K с единственным критическим простым числом 2. Соответствующие эллиптические кривые задаются уравнениями:

$$y^2 = x(x^2 - a), a = \pm 1, \pm 2, \pm 4, \pm 8.$$

$$y^2 = x(x - a\alpha)(x - a\alpha'), a = \pm 1, \pm 2; \alpha = 1 + \sqrt{-1}, 1 + \sqrt{2}, (1 + \sqrt{2})^2, 2 + \sqrt{2}.$$

Можно ли рассчитывать на то, чтобы получить достаточно полную классификацию полей алгебраических функций одной переменной по их множествам критических простых дивизоров?

Простейший случай, когда род g поля K равен 0, говорит в пользу такого предположения. В этом случае, как известно, поле функций имеет в качестве модели кривую, задаваемую уравнением

$$ax^2 + by^2 = 1.$$

Эта модель определяется полем K однозначно с точностью до про-

ективного преобразования. Классификация полей совпадает, следовательно, с классификацией тройничных квадратичных форм с точностью до линейной эквивалентности. Из теории квадратичных форм следует, что, во всяком случае, при $k=R$, поле рода 0 однозначно определяется множеством своих критических простых чисел. Совокупность всех полей, критические простые числа которых содержатся в заданном множестве $S=\{p_1, \ldots, p_m\}$ образуют абелеву группу, изоморфную прямому произведению m групп $\mathbf{Z}_2$.

Второй фундаментальной теоремой теории алгебраических чисел, связанной с критическими простыми числами, является теорема Минковского, согласно которой всякое конечное расширение K/R поля рациональных чисел R имеет хотя бы одно критическое простое число. Было бы очень интересно выяснить в какой мере этот результат обобщается на произвольные алгебраические многообразия. Вопрос в случае полей алгебраических функций одной переменной заключается в следующем:

Существуют ли поля алгебраических функций одной переменной с полем констант R, не имеющие критических простых чисел и отличные от поля рациональных функций $R(x)$?

Ответ на этот вопрос мне неизвестен. Для расширений произвольного поля алгебраических чисел k вопрос решается отрицательно. Больше того, существуют такие поля K алгебраических функций одной переменной, имеющие полем констант поле R, что для некоторого поля алгебраических чисел k поле $K \cdot k$ не имеет критических простых дивизоров над k. Примером является поле $R(x, y)$, где x и y связаны соотношением

$$x^3+y^3=1,$$

если за k взять поле $$R(\sqrt[3]{2}, \sqrt[4]{-3}).$$

Можно получить более тонкие инварианты, чем только совокупность критических простых дивизоров, если вернуться от рассмотрения полей к рассмотрению отдельных моделей. Для описания этой ситуации удобно пользоваться языком схем [14]. Пусть K/k — расширение конечного типа поля алгебраических чисел k, а X — схема над кольцом целых чисел σ поля k. Предположим, что схема X — целая, плоская и собственная над σ, что K является ее полем рациональных функций и что схема $X \otimes k$ (являющаяся алгебраическим многообразием над k), не особая. Понятие критического простого дивизора $\mathfrak{p}$ схемы X определяется, как и раньше, путем рассмотрения схемы $X/\mathfrak{p}$. Но в этом случае критическим простым дивизорам $\mathfrak{p}_i$ можно приписать некоторые кратности r_i и определить дискриминант схемы $D(X)=\prod \mathfrak{p}_i^{r_i}$. Не останавливаясь на том, как определяются числа r_i, скажу только, что это определение основывается на введенном Э. Келером [15] понятий дифференты (которая в этих терминах является замкнутой подсхемой схемы X), так же, как определение дискриминанта поля алгебраических чисел основывается на определении его дифференты.

Если поле K/k имеет степень трансцендентности 1, т. е. $X \otimes k$ является алгебраической кривой, то род поля K/k назовем родом схемы X. Имеет место следующий результат:

Существует только конечное число схем X заданного рода $g \neq 1$ (или рода 1 и заданного показателя d) с заданным дискриминантом.

Понятие дискриминанта схемы связано, повидимому, и с другими вопросами арифметики алгебраических многообразий. Например А. Нерон [16] доказал, что для $g=1$ среди всех схем X, соответствующих одному и тому же полю K, имеется одна, в определенном смысле минимальная. Можно показать, что эта схема обладает минимальным дискриминантом. Если, как можно надеяться, теорема о существовании минимальной модели имеет место для случая произвольного рода $g>0$, то эта минимальная модель будет также обладать минимальным дискриминантом.

§ 5. Сравнение с алгебраической геометрией

Вопросы об алгебраических кривых над полем алгебраических чисел k интересно сравнить с аналогичными вопросами для случая, когда k — поле алгебраических функций одной переменной, подобно тому, как мы уже это делали в этом докладе в связи с другими вопросами теории алгебраических чисел.

Пусть k — поле алгебраических функций одной переменной с алгебраически замкнутым полем коэффициентов k_0, являющееся полем рациональных функций на полной алгебраической кривой T. Рассмотрим, аналогично тому, как мы поступали в § 4 поле алгебраических функций одной переменной K с полем коэффициентов k.

Так как K имеет над k_0 степень трансцендентности 2, то оно является полем рациональных функций на алгебраической поверхности X, которую можно считать неособой. Вложение поля k в K определяет рациональное отображение $f\colon X \to T$. Пользуясь известными приемамы теории алгебраических поверхностей, можно свести вопрос к случаю, когда отображение f регулярно. Таким образом, аналогом той ситуации, которая обсуждалась в § 4 является регулярное отображение неособой алгебраической поверхности X на неособую алгебраическую кривую T. Для всех точек t кривой T, кроме, быть может, конечного числа их, прообраз $f^{-1}(t)$ будет неособой алгебраической кривой одного и того же рода g, равного роду поля K/k. Те точки, прообразы которых являются кривыми с особенностями и служат аналогом критических простых дивизоров.

Мы будем называть их критическими точками.

Говоря другими словами, мы имеем расслоение поверхности X на семейство алгебраических кривых $f^{-1}(t)$. Критические точки соответствуют вырождающимся слоям, которые являются кривыми с особыми точками.

Обратимся к вопросу о полях без критических точек.

Таким полям будут соответствовать расслоения без вырожденных слоев. В нашем случае такие расслоения действительно существуют. Примером их являются прямые произведения $X=F\times T$ базисной кривой T на произвольную кривую F. Более общий класс примеров дают «постоянные» расслоения, у которых все слои бирегулярно эквивалентны друг другу. Эти расслоения являются алгебраическими расслоенными пространствами, локально тривиальными при $k_0=C$ и ло-

кально изотривиальными [17] при произвольном k_0. Вопрос о полях без критических точек приобретает теперь следующую форму:

Совпадает ли всякое расслоение без вырожденных слоев, база T которого является проективной прямой, с «постоянным расслоением»?

Этому вопросу можно придать другую форму, если рассмотреть многообразие модулей M_g алгебраических кривых рода g. Сопоставляя слою $f^{-1}(t)$ $(t \in T)$ точку, изображающую эту кривую в многообразии M_g, мы определим регулярное отображение проективной прямой T в M_g. Вопрос заключается в том, чтобы выяснить, существуют ли такие отображения, кроме отображений в точку. Их заведомо нет, если многообразие M_g аффинное, как это имеет место при $g=2$ (1).

Перейдем к рассмотрению расслоений, содержащих вырожденные слои. Здесь удобно (и, возможно, не теряя общности), ограничиться рассмотрением минимальных расслоений, т. е. таких расслоений $X \xrightarrow{f} T$, что не существует расслоения $X' \xrightarrow{f'} T$ и такого регулярного бирационального отображений g: $X \to X'$ что треугольник

$$\begin{tikzcd} X \arrow[rr, dashed, "g"] \arrow[dr, "f"'] & & X' \arrow[dl, "f'"] \\ & T & \end{tikzcd}$$

коммутативен.

Вопросу о полях с заданными критическими точками может быть придана следующая форма:

Конечно ли число расслоений на кривые заданного рода $g \neq 1$, если фиксирована базисная кривая T и множество S критических точек расслоения?

В ряде частных случаев (например, если все слои гипер-эллиптические) можно показать, что этот вопрос имеет положительный ответ. Доказательство в общем случае должно быть, повидимому, значительно труднее доказательства аналогичного факта для кривых над полями алгебраических чисел, аналогично тому, как конечность числа расширений с заданными точками ветвления поля алгебраических функций доказывается гораздо сложнее [18], чем теорема Эрмита в теории алгебраических чисел. В связи с этим было бы интересно исследовать этот вопрос в случае $k_0=C$, пользуясь топологическими методами.

В § 1 мы видели, что при $k_0=C$ классификация конечных расширений поля алгебраических функций может быть сведена к топологической классификации накрытий.

Связи, которые существуют между теорией регулярных отображений алгебраических поверхностей на алгебраические кривые и соответствующей топологической задачей — теорией дифференцируемых отображений четырехмерных дифференцируемых многообразий на двумерные — по-видимому, совсем не исследованы. Здесь возникают три вопроса:

Какова классификация дифференцируемых отображений четырехмерных многообразий на двумерные?

---

(1) D. Mumford сообщил мне, что K. Kodaira построил пример не«постоянного» расслоения без вырожденных слоев с базой, отличной от проективной прямой.

Какие дифференцируемые отображения четырехмерного многообразия на алгебраическую кривую реализуются как регулярные отображения алгебраических поверхностей на эту кривую?

Могут ли два различных регулярных отображения алгебраической поверхности на алгебраическую кривую быть диффеоморфными?

Последний вопрос аналогичен тому, который Г. Грауерт [19] решил в положительную сторону для расслоений с векторным слоем над многообразиями Штейна.

Заключение

В этом докладе я не ставил своей целью дать сколько-нибудь полный обзор результатов, полученных в последнее время в теории алгебраических чисел. Здесь рассмотрен только один круг идей в теории полей алгебраических чисел и собраны примыкающие к этим идеям результаты и проблемы в самой теории алгебраических чисел и в близких к ней областях.

ЛИТЕРАТУРА

[1]. Grothendieck, A., Géométrie formelle et géométrie algébrique. *Séminaire Bourbaki*, Mai, 1959, 182, 1–28.

[2]. Шафаревич, И. Р., О p-расширениях. *Математический сборник*, 20, 2 (1947), 351–363.

[3]. Kawada, J., On the structure of the Galois group of some infinite extensions. I. *J. Fac. Sci. Univ. Tokyo*, Sec. I, 7, 1 (1954), 1–18.

[4]. Фаддеев, Д. К., и Скопин, А. И., К доказательству одной теоремы Кавада. *Доклады АН СССР*, 127, 3 (1959), 529–530.

[5]. Zassenhaus, H., *Lehrbuch der Gruppentheorie*. Leipzig, 1937.

[6]. Brauer, R., Über die Konstruktion der Schiefkörper, die von endlichem Rang in Bezug auf ein gegebenes Zentrum sind. *J. reine angew. Math.*, 168, 1 (1932), 44–64.

[7]. Демушкин, С. П., Группа максимального p-расширения локального поля. *Известия Акад. наук СССР, серия матем.*, 25, 3, 329–346.

[8]. Koch, H., Über galoissche Erweiterungen p-adischer Zahlkörper. *J. reine angew. Math.*, 209, 1–2 (1962), 9–12.

[9]. Гаусс, К. Ф., *Арифметические исследования* в. (К. Ф. Гаусс), Труды по теории чисел. Москва, 1959.

[10]. Scholz, A., Zwei Bemerkungen zum Klassenkörperturm. *J. reine angew. Math.*, 161, 4 (1929), 201–208.

[11]. Fröhlich, A., A note on the class field tower. *Quart. J. Math. (Oxford)*, 5, 5 (1954), 141–144.

[12]. Shimura, G., Reduction of algebraic varieties with respect to a discrete valuation of the basic field. *Amer. J. Math.*, 77, 1 (1955), 134–176.

[13]. Шафаревич, И. Р., Показатели эллиптических кривых. *Доклады АН СССР*, 114, 4 (1957), 714–716.

[14]. Grothendieck, A. & Dieudonné, J., *Eléments de géométrie algébrique*. I. *Publ. Math., (I.H.E.S.), Paris*, 8.

[15]. Kähler, E., Geometria arithmetica. *Ann. Mat. Pura Appl.*, Ser. IV, 45 (1958), 1–399.

[16]. Néron, A., Modèles p-minimaux des variétés abéliennes. *Séminaire Bourbaki*, Décembre, 1961, 227, 1–16.

[17]. Serre, J.-P., Espaces fibrés algébriques. *Séminaire C. Chevalley*. École Normale Supérieure, Exp. 1 (1958), 1–37.

[18]. Lang, S. & Serre, J.-P, Sur les revêtements non ramifiés des variétés algébriques. *Amer. J. Math.*, 79, 2 (1957), 319–330.

[19]. Grauert, H., Analytische Faserungen über holomorph-vollständige Räume. *Math. Ann.*, 135 (1958), 263, 273.

DISCONTINUOUS GROUPS AND HARMONIC ANALYSIS

By ATLE SELBERG

1. We shall in the following give a brief account of certain results and problems that arise in connection with discontinuous groups operating on certain types of spaces and the functions that are associated with these in a natural way.

The analytic automorphic forms of one or more complex variables have long been known, in recent times notably through the work of Hecke and Siegel, to be intimately connected with questions from the theory of numbers. Maass [1] introduced, in order to extend certain investigations of Hecke, automorphic functions characterized by being eigenfunctions of certain differential operators. Later the emphasis was shifted to the study of certain rings of integral operators, and in particular to a general formula which I have called the "trace formula".

The principal features of this theory carries over to cases where there exist no differential operators, but we shall here, apart from the earlier part of section 2, be concerned only with the groups operating on and functions defined on symmetric riemannian spaces (in the sense of E. Cartan). Section 2 is devoted to a short summary of this theory for the case of groups with compact fundamental domain. I have here little to add to what has been said earlier by myself and others.(1) Section 3, the main part of this exposition, deals with certain of the problems introduced by noncompactness, especially the problem of analytic continuation of Eisenstein series. The last section concerns applications to the theory of Hecke operators.

I have not tried to give explicit examples nor to go into the number-theoretical aspects and applications of results stated as this would have made this exposition too long.

2. Let G be a locally compact group with bi-invariant Haar measure and S a space that is defined as the quotient space G/K where K is some compact subgroup of G. We shall denote the elements of G by g and points in the space S by x, y. We shall consider scalar functions $f(x)$ and let g operate on the functions in the manner that $f(x) \to f(gx)$; we might more generally consider vector valued functions and $f(x) \to \sigma(g,x) f(gx)$, where $\sigma(g,x)$ is a scalar or matrix valued multiplier satisfying the relation $\sigma(g_2,x)\sigma(g_1,g_2x) = \sigma(g_1g_2,x)$, but shall not do so here except in connection with what we call the finite case of the trace formula.

We consider a ring of operators of the form

$$\int_S k(x,y)\, f(y)\, dy, \tag{2.1}$$

where dy enotes a measure on S that is invariant under the action of G, and where the kernel is such that the operator commutes with the action of G on f, this means that

(1) Selberg [8] and [9], Tamagawa [12], Godement [2].

$$k(x,y)=k(gx,gy) \tag{2.2}$$

for all g. We assume that this ring is commutative, an assumption which can be verified for instance if S is a symmetric riemannian space and G the connected component of its group of isometries. We shall consider functions of x which are eigenfunctions of all the operators in the ring. We are particularly interested in eigenfunctions $\omega(x,y)$ which satisfy (2.2) and for which $\omega(y,y)=1$. From an arbitrary eigenfunction $f(x)$, we can obtain one of this form by defining for $f(y)\neq 0$,

$$\omega(x,y)=\frac{1}{f(y)}\int_{K_y} f(\varrho x)\,d\varrho, \tag{2.3}$$

where K_y is the compact subgroup of G that leaves y fixed and $d\varrho$ is an invariant measure on K_y such that the total measure of K_y is 1. Every eigenfunction $f(x)$ is in this sense associated with a unique $\omega(x,y)$, and its eigenvalue with respect to the operator (2.1) is therefore the same as that of $\omega(x,y)$, which is

$$h(\omega)=\int_S k(x,y)\,\omega(y,x)\,dy. \tag{2.4}$$

Instead of using ω as a parameter here, we can if we have a set of operators D_i that commute with our ring and with the action of G on S, and which is such that each $\omega(x,y)$ is uniquely determined by the set of eigenvalues $\lambda=\{\lambda_i\}$ where λ_i is the eigenvalue of ω with respect to D_i, use this set of numbers λ as a parameter and write $h(\lambda)$ instead of $h(\omega)$. If S is a riemannian space a basis for the ring of differential operators that commute with the action of G will provide such a set D_i.

We now assume that we have a discrete subgroup Γ of G, and we consider now functions defined of $\Gamma\backslash S$, that is functions for which $f(x)=f(\gamma x)$ for $\gamma\in\Gamma$. By a fundamental domain D we understand as usual a set of points such that $S=\sum_{\gamma\in\Gamma}\gamma D$. For our new class of functions f, the operator (2.1) takes the form

$$\int_D K(x,y)\,f(y)\,dy, \tag{2.5}$$

where

$$K(x,y)=\sum_{\gamma\in\Gamma} k(x,\gamma y). \tag{2.6}$$

We shall assume throughout this section that $\Gamma\backslash S$ or D is compact, this actually also means that D may be chosen so as to lie entirely in some compact subset of S.

The trace formula is now simply the result that we formally obtain by evaluating the trace of the operator (2.5) on the one hand as

$$\int_D K(x,x)\,dx,$$

using the expression for K and the properties of k, this can be transformed into a series of the form

$$\sum_{\{\gamma\}_\Gamma} \mu(G_\gamma/\Gamma_\gamma)\, g(\{\gamma\}_G), \tag{2.7}$$

where Γ_γ and G_γ mean the centralizer of γ in Γ and G respectively, $\mu(G_\gamma/\Gamma_\gamma)$ is a measure of the quotient G_γ/Γ_γ, the symbol $\{\gamma\}_G$ denotes the conjugacy class of γ in G, g depends on k and the argument $\{y\}_G$, and the symbol $\{\gamma\}_\Gamma$ denotes conjugacy class of γ in Γ and we sum over one representative from each conjugacy class. On the other hand we have also that the trace of the operator (2.5) is given by

$$\sum_\omega h(\omega) \quad \text{or} \quad \sum_\lambda h(\lambda), \tag{2.8}$$

where h is given by (2.4) and the summation extended over the set of ω's (or λ's) for which there exist eigenfunctions, each ω counted with the proper multiplicity.

The trace formula

$$\sum_\omega h(\omega) = \sum_{\{\gamma\}_\Gamma} \mu(G_\gamma/\Gamma_\gamma)\, g(\{\gamma\}_G) \tag{2.9}$$

will be valid under certain conditions that shall not be given here.

Two obvious ways in which this identity may be used are to get information about the distribution of the terms occurring on the right-hand or the left-hand side. To obtain the finer results one would use methods resembling those used in analytic number theory. Actually for some of the simpler cases of symmetric riemannian spaces of negative curvature (2.9) turns out to be structurally strikingly similar to the socalled "exact formulas" in the theory of primes, so much so that one can actually from it construct a kind of zeta function associated with the group Γ, which in its properties (Euler product, functional equation, and location of trivial and nontrivial zeros) resembles the usual zeta or L-functions.

In general the series occurring on both sides of (2.9) are infinite, if G is compact and therefore Γ finite, the series on the right-hand side of (2.9) is of course finite, and the series on the left-hand side would be finite if k is a finite linear combination of ω's. For noncompact G there may however also be exceptions, if there are among the ω's functions that would qualify as k in the trace formula. This we shall refer to as a "finite case" of the trace formula. The most notable instance is connected with the theory of analytic automorphic forms of one or several complex variables. This is actually a case where we have as multiplier $\sigma(g,x)$ a power of the jacobi determinant $j_g(z)$ of the analytic transformation $z \to gz$. In this case, using the proper k, which can be simply expressed in terms of the Bergmann kernel function if S is a bounded homogeneous domain, the left-hand side of (2.9) becomes simply $N(r,\Gamma)$, the number of linearly independent regular analytic automorphic forms of weight r (that is, functions which transform in the way $f(z) = (j_\gamma(z))^r f(\gamma z)$ for γ in Γ); on the right-hand side of (2.9) all terms g where γ is not a transformation with fixpoint in S will vanish, leaving a finite sum. The terms with fixpoints yield polynomials in r whose degree equals the complex dimension of the fixpoint set of γ, multiplied with the rth power of a certain root of unity (the value of $j_\gamma(z)$ at a fix point).(1)

(1) Selberg [9].

More generally if we instead of $(j_g(z))^r$ have a matrix multiplier $\sigma(g,z)$ which is analytic in z, the same approach works. Recently Langlands(1) has given the explicit expression for the number of linearly independent analytic forms $N(\sigma, \Gamma)$ by evaluating the relevant $g(\{\gamma\}_G)$'s in closed form. As the general expression for $g(\{\gamma\}_G)$ involves considerable Lie group terminology we shall not give it here.

3. If we give up the assumption of the compactness of D, the situation becomes much more complicated. In order to get a sensible class of groups we have to require at least that the measure of D be finite. In general when D is not compact the trace of our integral operator (2.5) does not exist, and we have beside the discrete spectrum of eigenfunctions, one or more continous spectra, these may be of various dimensions. The general idea one would try to follow in order to re-establish a "trace-formula" in some modified form for this case would be first, to obtain enough information about the continuous spectra, then to remove their contribution to the kernels of the operators (2.5) so that one would get a class of operators that only retain the discrete part of the spectrum, and these eigenfunctions with unchanged eigenvalues. Finally one would then try to evaluate the trace of these new operators in two different ways.

From now on we shall consider only the case that S is a symmetric riemannian space which for simplicity may be assumed to have no compact factor spaces. The most serious obstacle to the carrying out of the idea sketched above is that, with the exception of the case of the hyperbolic plane (and of course for the product of a hyperbolic plane and euclidean or compact symmetric spaces) it is not known what the conditions of the noncompactness but finite measure or volume of D imply about Γ. It is not in general known if such a group is necessarily finitely generated. If by a canonically constructed fundamental domain we understand one obtained from a base point(2) by including in D any point whose geodesic distance from the base point is less than from any point equivalent under Γ, then it is not known whether such a fundamental domain necessarily is bounded by a finite number of sides. Finally it is not known whether such a group necessarily always contains unipotent elements.

Although in certain cases(3) it has been possible to draw definite conclusions about the nature of the noncompact parts of the fundamental domain D, the "cusps" as we shall call them in analogy with the terminology used for the hyperbolic plane, and the subgroups of Γ that describe these, by making rather weak assumptions in addition to the finiteness of the volume of D, it seeems reasonable at present to concentrate the efforts of handling the difficulties introduced by noncompactness, in two directions. The more general approach would be to make definite assumptions about the nature of the cusps and the subgroups of Γ that describes them, namely that these be of the same nature as those encountered for the arithmetic subgroups of algebraic groups in the sense of Borel–Harish Chandra.(4) There are reasons

---

(1) Communicated at a conference on automorphic functions at Stanford in June 1962.

(2) Assumed not to be a fixpoint of any element in Γ.

(3) See Piatetsky–Shapiro [6].

(4) Borel–Harish Chandra [1].

to believe that perhaps all groups Γ with finite volume of D would fall in this category.

The other approach would be to make a definite assumption about the whole group Γ, namely that Γ by a suitable representation of G as an algebraic group is commensurate with an arithmetic subgroup. (We call two groups commensurate if their intersection has finite index in both.) This more restricted approach could be justified by the fact that for the higher dimensional symmetric spaces the arithmetic subgroups are the only types of groups with finite volume of D that have been constructed, if we exclude certain obvious exceptions for certain product spaces.

We shall first describe what can be done about the continuous spectra with the first approach, and assume that all cusps are of a particularly simple kind, and that there are only a finite number of inequivalent ones.(1)

Let us assume that S whose points we shall denote by z (not to be thought of here as denoting one or more complex variables) can be represented by coordinates $z=(y,u,x)$, where y is real positive, u represents a point in a symmetric space S_u (possibly euclidean or with euclidean factors and possibly even zero dimensional) of lower dimension than S, and X represents an element of a certain unipotent subgroup N_x of G. Further let G contain a subgroup G_∞ which is transitive on S. G_∞ contains N_x, the elements n of N_x act on z in the way that (y,u) is left unchanged while $x \to nx$, N_x is furthermore a normal subgroup of G_∞. An element of G_∞ acts on z in the way that y is multiplied by a positive constant and u undergoes a transformation which belongs to a group G_u of isometries of S_u. The restriction of G_∞ to the (y,u) space can therefore be written as $G_{u,y} = G_u \cdot G_y$, where G_y is the multiplicative group of the positive real numbers. The invariant volume element of S becomes

$$dz = c_1 \frac{dy}{y^\alpha}\, du\, dx, \tag{3.1}$$

where we use c_i $(i=1,2,\ldots)$ to denote positive real constants, and α is a constant which we assume to be >1, by writing y instead of $y^{1/(\alpha-1)}$ we can always attain that $\alpha=2$. Further du is an invariant volume element on S_u and dx the Haar measure on the unipotent group N_x. The metric form on S would take the form

$$ds^2 = c_2 \frac{dy^2}{y^2} + c_3\, ds_u^2 + Q_{y,u,x}(dx), \tag{3.2}$$

where ds_u is the invariant line element in S_u, $Q_{y,ux}(dx)$ is a positive definite quadratic form in the dx, whose coefficients depend on y, u and x, whose determinant is c_4/y^2, and which furthermore has the property that for u and x in some compact region of S_u and N_x, its coefficients tend uniformly to zero as $y \to \infty$.

We now denote the intersection $\Gamma \cap G_\infty$ by Γ_∞, further $\Gamma \cap N_x$ by Γ_x. The restriction of Γ_∞ to the (y,u) space shall be a discrete subgroup of G_u (this means that Γ_∞ does not change y) which we denote by Γ_u. We assume

(1) The latter part of this assumption can in many cases be shown to follow from the first.

that $\Gamma_x \backslash N_x$ is compact, and also that $\Gamma_u \backslash G_u$ is compact. We say then that Γ_∞ defines a "point-like" cusp at $y=\infty$. It can be seen that if we define S_{y_0} as the part of S with $y \geqslant y_0$, then for y_0 sufficiently large D can be chosen such that the intersection of D and S_{y_0} is simply a fundamental domain of Γ_∞ in S_{y_0}. Such a cusp, when it is part of a canonically constructed fundamental domain, contains exactly one geodesic ray from the base point which remains inside the cusp. We shall assume that Γ has only a finite number of nonequivalent cusps of this nature, and that what remains of D after we "cut off the cusp," that is, remove the part with $y > y_0$ for y_0 sufficiently large, for all these cusps, is compact. That the cut-off part has finite volume is obvious from the form of the volume element. The cut-off part of the cusp could be given as $y > y_0$, u in D_u and x in D_x where D_u is a fundamental domain of Γ_u in S_u and D_x a fundamental domain of Γ_x in N_x.

If we have defined in S a function $f(z)$ which is invariant under Γ_∞ we may form

$$\frac{1}{m(D_x)} \int_{D_x} f(nz)\, dn, \tag{3.3}$$

where $m(D_x)$ is the measure of D_x. We shall refer to this expression, which will be a function of (y,u) which is invariant under Γ_u, as the "constant term" of f with regard to the cusp. I use this name because in the (in application frequently occurring) case that N_x is abelian this is actually the constant term in the fourier expansion of f with regard to x. If the constant term vanishes we say that f is a cusp form with respect to this cusp.

If f is an eigenfunction of the operators (2.1), we can show, that if f is a cusp form and $f = O(y^k)$ for some k as $y \to \infty$, then this holds for all k, so that f tends to zero faster than any negative power of y. Also, if the eigenvalues λ to which f belongs for a basis of the invariant differential operators are restricted to a compact region, we can show that f does not attain any maximum in the region $y > y_0$ for some fixed y_0 which depends only on this compact region to which the eigenvalues are restricted and on Γ_∞.

If s is a complex variable and $v(u)$ is invariant under Γ_u and also an eigenfunction of the invariant operators on the space S_u, then we have that

$$f_{s,v}(z) = y^s v(u) \tag{3.4}$$

is an eigenfunction of our class of invariant operators on S. If we form for $s = \sigma + it$, $\sigma > 1$ the series

$$E(z, s, v) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} f_{s,v}(\gamma z), \tag{3.5}$$

this series is seen to converge absolutely and to represent an eigenfunction which is invariant under Γ, we call the function defined by (3.5) the Eisenstein series corresponding to the cusp and to the function v. The continuous spectra for our problem are in this case obtained by analytic continuation of this series as a function of s for $s = \frac{1}{2} + it$. Under the assumptions we made earlier about Γ it can be proved that this analytic continuation is always possible, that the $E(z,s,v)$ for the various cusps are meromorphic functions of s in the whole s-plane (with location of singularities independent of z), that they are regular for $\sigma \geqslant \frac{1}{2}$ except possibly for a finite number of

poles of order 1 on the line segment $\frac{1}{2}<s\leqslant 1$. The residue of $E(z,s,v)$ at such a pole, gives an eigenfunction belonging to the discrete spectrum which is not a cusp form (actually all such eigenfunctions arise in this way). We also obtain functional equations that relate the set of Eisenstein series which corresponds to the same eigenvalues (or to the same ω) for variable s, to another (or possibly the same) such set with $1-s$ instead of s as argument.

We shall in the following give some indications of the ideas on which the proofs of these results are based. For simplicity we assume that D has just one cusp, and furthermore that v is such that if for another function v' we have that $f_{s,v}(z)$ and $f_{s,v'}(z)$ correspond to the same ω_s for variable s then v' is a constant multiple of v.

If we apply the operation (3.3) to $E(z,s,v)$ we can show that the "constant term" must have the form

$$y^s v(u)+\varphi(s)y^{1-s}w(u), \tag{3.6}$$

where $w(u)$ is a function of the same type as v and such that $y^{1-s}w(u)$ belongs to the same ω_s as $y^s v(u)$, and $\varphi(s)=\varphi_v(s)$ is a certain function of s which is regular for $\sigma>1$. If we form $E(z,s,w)$ we can show that its "constant term" has the form

$$y^s w(u)+\overline{\varphi(\bar{s})}\,y^{1-s}v(u) \tag{3.6'}$$

i
f v and w are both normed such that the integral of the square of their absolute value over D_u is 1. There are two possible cases to consider. Either $w(u)$ is apart from a constant factor identical to $v(u)$, or $w(u)$ and $v(u)$ are orthogonal so that

$$\int_{D_u} v\bar{w}\,du=0.$$

The latter case is the simpler so we shall consider it first here. We write for z in D

$$\tilde{E}(z,s,v)=\begin{cases} E(z,s,v) & \text{for} \quad y<y_0, \\ E(z,s,v)-y^s v(u) & \text{for} \quad y>y_0, \end{cases} \tag{3.7}$$

and similarly

$$\tilde{\tilde{E}}(z,s,v)=\begin{cases} E(z,s,v) & \text{for} \quad y<y_0, \\ E(z,s,v)-y^s v(u)-\varphi(s)\,y^{1-s}w(u) & \text{for} \quad y>y_0. \end{cases} \tag{3.7'}$$

It is possible to show that for a $s=\sigma+it$ with $\sigma<\frac{1}{2}$ for which $E(z,s,v)$ is analytic we have

$$\int_D \left|\frac{d^n}{ds^n}\tilde{E}(z,s,v)\right|^2 dz=c_5 2^{-2n}\frac{d^{2n}}{d\sigma^{2n}}\left(\frac{y_0^{2\sigma-1}}{2\sigma-1}\right), \tag{3.8}$$

where c_5 is a constant, independent of v, but dependent only on S and Γ_∞. From this it is not hard to conclude that $E(z,s,v)$ is regular in the half plane $\sigma>\frac{1}{2}$. To obtain results going further we need the formula

$$\int_D |\tilde{E}(z, s, v)|^2 dz = c_5 \frac{y_0^{2\sigma-1} - |\varphi(s)|^2 y_0^{1-2\sigma}}{2\sigma - 1}, \tag{3.9}$$

valid wherever $E(z,s,v)$ is analytic, thus in particular for $\sigma > \frac{1}{2}$. From (3.9) it is seen that $|\varphi(s)| < y_0^{2\sigma-1}$ for $\sigma > \frac{1}{2}$. We next show that $|\varphi(s)| \to 1$ as $\sigma \to \frac{1}{2}$. If this were not the case, we denote the right-hand side of (3.9) by $\Omega(s)$ and consider

$$\frac{E(z, s, v)}{\sqrt{\Omega(s)}} \tag{3.10}$$

as $\sigma \to \frac{1}{2}$. If there is a sequence of $s_j = \sigma_j + it_j$ with t_j bounded and $\sigma_j \to \frac{1}{2}$, such that $\Omega(s_j) \to \infty$ for $j \to \infty$, then we can show that we could take out a subsequence of the s_j such that the expression (3.10) would converge towards a square-integrable eigenfunction which would be a cusp form. This easily leads to a contradiction since such a function is orthogonal to all Eisenstein series. Thus $\lim_{\sigma \to \frac{1}{2}} |\varphi(s)| = 1$ and we see that $\varphi(s)$ can be continued over the line $\sigma = \frac{1}{2}$ and satisfies the functional equation

$$\varphi(s)\overline{\varphi(1-\bar{s})} = 1, \tag{3.11}$$

$\Omega(s)$ therefore tends to a finite limit as the point s approaches a point on the line $\sigma = \frac{1}{2}$ from the right, and it is then not hard to show that also $E(z,s,v)$ has continuous boundary values as $\sigma \to \frac{1}{2}$, and can be continued across the line $\sigma = \frac{1}{2}$ by means of the functional equation

$$E(z,s,v) = \varphi(s) E(z, 1-s, w). \tag{3.12}$$

Thus E in this case is regular for $\sigma \geqslant \frac{1}{2}$ and meromorphic everywhere.

In the other case where w apart from a constant factor is the same function as v, we may take $w = v$. In this case the formula that takes the place of (3.8) becomes quite complicated as the function φ enters, however it is possible to obtain an inequality which gives analyticity for $\sigma > \frac{1}{2}$, with the exception of the real interval $\frac{1}{2} < s \leqslant 1$. The expression on the right-hand side of (3.9) also acquires some new terms but our argument about the line $\sigma = \frac{1}{2}$ goes through much as before; we still have to deal with the real interval $\frac{1}{2} < s \leqslant 1$, and find that we may possibly have a finite number of simple poles here. Apart from this (3.11) and (3.12) remain valid.

If we have several v's such that $f_{s,v}(z)$ belongs to the same ω_s, the same methods go through. Instead of (3.12) we have an equation describing how the system of the $E(z,s,v)$ is related to a system $E(z, 1-s, w)$ by means of a matrix which becomes unitary for $s = \frac{1}{2} + it$. Similarly if we have more than one cusp, we then consider the set of Eisenstein series formed with respect to all inequivalent cusps, and for the system obtained by considering those belonging to the same ω_s, the corresponding results still hold.

I may mention an open problem in connection with the possibility of the simple poles on the interval $\frac{1}{2} < s \leqslant 1$, for $v = 1$ identically there is always a pole at $s = 1$; for the hyperbolic plane we can construct groups for which the Eisenstein series with $v = 1$ (which is the only one existing in that case) has arbitrarily many poles between $\frac{1}{2}$ and 1. For the arithmetically defined groups however as the modular group and its congruence subgroups only

the pole at 1 is present. It would be interesting to know whether in higher dimensional cases, excluding again certain obvious exceptions where S is a product space with a hyperbolic plane as a factor, there may exist other poles than the one at $s=1$ for $v=1$ identically.

The $E(z, \frac{1}{2}+it, v)$ obtained from the various cusps and v's are functions in the continuous spectrum. It is not hard to show under the assumptions we have made about Γ that if the kernel k of our integral operator satisfies certain conditions, we can form an auxilliary kernel $H(z_1, z_2)$ from the Eisenstein series for $s=\frac{1}{2}+it$ and the functions $h(\omega_{s,v})$ in such a way that the integral operator with kernel $K-H$ retains only the discrete part of the spectrum and with unchanged eigenvalues and actually has a finite trace. The evaluation of the trace of this modified operator, which would then yield the modified trace-formula, has only been carried out explicitly in some special cases.

If we give up the requirement that $\Gamma_u \backslash S_u$ or D_u be compact for the cusps, but only require finite volume, we could still define Eisenstein series in the same manner as before if v is a square-integrable eigenfunction invariant under Γ_u, however, the proof sketched out above does not seem to carry over too easily to this case. But it would not be sufficient to consider only these Eisenstein series, which would give rise to the one-dimensional continuous spectra. In general there would be in a canonically constructed fundamental domain not only a finite number of isolated geodesics issuing from the base point which remain inside D. Corresponding to such a family, we could assume that there is a parametrization of the space S in coordinates $(y_1, y_2, \ldots, y_r, u, x)$ with $r>1$ and otherwise with requirements corresponding to those we made about the group G_∞ and Γ_∞ before. We could then define series

$$\sum_{\gamma \in \Gamma_\infty \backslash \Gamma} f_{s_1, s_2 \ldots s_r; v}(\gamma z), \tag{3.13}$$

where

$$f_{s_1, s_2 \ldots s_r, v}(z) = y_1^{s_1} y_2^{s_2} \ldots y_r^{s_r} v(u). \tag{3.14}$$

These would be Eisenstein series that are dirichletseries in r complex variables. It seems very difficult to attack the problem of continuation of these without making explicit assumptions about the whole group Γ rather than just parts of it. There is however some possibility that one might take some advantage in the case that $r>1$ of the general theory of functions of several complex variables and so not have to continue the series (3.13) into the region where the real parts of several of the s become small, but rather just would need to establish continuation where one of the real parts of the s_i becomes small and the others stay larger than some suitable large constant. The functional equations one thus could establish might then provide enough information to deduce the rest from the general theory. Nothing has been done along these lines so far however.

It might seem well justified however, at least at present, to try to attack the problem of the analytic continuation of the Eisenstein series under the assumption that by a suitable representation of G as an algebraic group Γ is commensurate to an arithmetic subgroup in the sense of Borel–Harish Chandra.

Only for a few special cases however has this continuation actually been carried out. In the case that G is the group usually denoted by $Sl(n, R)$ and

S the space of symmetric positive definite matrices Y with determinant 1, and Γ the group of all n by n integral matrices with determinant 1, it has been possible to effect the continuation of all the Eisenstein series. We shall here give the properties of the series that is dependent on the largest number of complex variables $s_1, s_2, \ldots, s_{n-1}$(1).

Let $\overset{n}{Y}$ denote a point in S and in general $\overset{\nu}{Y}$ a ν by ν positive definite symmetric matrix. We call a sequence

$$\overset{n}{Y} \to \overset{n-1}{Y} \to \ldots \to \overset{1}{Y} \tag{3.15}$$

a descending chain, if each $\overset{\nu-1}{Y}$ is obtained from the preceding $\overset{\nu}{Y}$ by transforming it by a $\nu-1$ by ν integral matrix. If no $\overset{\nu}{Y}$ has determinant zero we call the chain nondegenerate, and we call it primitive if each of the $\nu-1$ by ν integral matrices can be completed to an integral ν by ν matrix of determinant 1. Two such descending chains are called equivalent if each pair of corresponding links are equivalent forms.

We may now consider the two series

$$\zeta_Y^*(s_1, s_2 \ldots s_{n-1}) = \sum{}^* \prod_{\nu=1}^{n-1} |\overset{\nu}{Y}|^{-s_\nu}, \tag{3.16}$$

and

$$\zeta_Y(s_1, s_2 \ldots s_{n-1}) = \sum \prod_{\nu=1}^{n-1} |\overset{\nu}{Y}|^{-s_\nu}, \tag{3.16'}$$

where in the first we sum over a full set of inequivalent primitive chains, and in the second case over a full set of inequivalent nondegenerate chains. If $s_\nu = \sigma_\nu + it_\nu$ we can show that both series converge absolutely for $\sigma_\nu > 1$; $\nu = 1, 2, \ldots, n-1$. Further one can show that

$$\zeta_Y = \zeta_Y^* \prod_{1 \leqslant \nu \leqslant \mu \leqslant n-1} \zeta(2 \sum_{\nu}^{\mu} s_j - (\mu - \nu + 1)), \tag{3.17}$$

where ζ is the Riemann zeta function. The analytic properties are simplest described in terms of a new set of variables $z_j (j = 1, 2 \ldots n)$, such that $s_j = z_{j+1} - z_j + \frac{1}{2}$. If we put $\zeta_Y(s_1 \ldots s_{n-1}) = Z_Y(z_1, z_2 \ldots z_n)$, we can show that the expression

$$\prod_{1 \leqslant i < j \leqslant n} (z_j - z_i - \tfrac{1}{2}) \cdot Z_Y(z_1 \ldots z_n) \tag{3.18}$$

is holomorphic everywhere in complex n-space and further that the expression

$$\pi^{-2 \sum_{j=1}^{n} j z_j} \prod_{1 \leqslant i < j \leqslant n} \Gamma(z_j - z_i + \tfrac{1}{2}) Z_Y(z_1, z_2 \ldots z_n) \tag{3.19}$$

is invariant under any permutation of the z_j. Although the functional equations (of which there in this case are many) and the analytic behavior are simplest described in terms of ζ_Y, it is ζ_Y^* which is the true Eisenstein series in this case and the $(n-1)$-dimensional continuous spectrum is obtained from ζ_Y^* for $s_j = \frac{1}{2} + it_j$.

(1) Also briefly described in Selberg [10].

It may be remarked in passing that it is possible to define series like ζ_Y also for integral indefinite nondegenerate forms. There is then a system of series (the number of which depends on the signature), these systems have analogous analytic properties.

In addition to ζ_Y we have in this case other Eisenstein series that define continuous spectra of lesser dimensions. These of course have a smaller number of functional equations.

R. P. Langlands has proved(1) that the Eisenstein series in n complex variables which can be defined for the symplectic modular group acting on the space of symmetric complex n by n matrices with positive definite imaginary part has analytic properties similar to those of ζ_Y^*.

However, in these cases where the continuous spectra are multidimensional, probably a much bigger step remains after one has determined the properties of these spectra in order to get to a trace-formula, than in the case spoken about earlier where only one dimensional continuous spectra occur.

Noncompactness, but with finite volume, of the fundamental domain creates also difficulties in connection with what we in the previous section called the finite case of the trace formula, although these difficulties are not as deep as those spoken of earlier. It is however still an open problem to determine the contribution of the cusps (or rather the subgroups of Γ that describes the cusps) to the number $N(r,\Gamma)$ of linearly independent automorphic forms of weight r. The only result of any generality that has been obtained in this direction can be described as follows:

Let us assume first that for each cusp it is possible to represent the bounded domain as a Siegel domain of the first kind(2) with complex coordinates $z_1 \dots z_n$ which we consider as components of a column vector $z = x + iy$, such that the imaginary part y lies in a positivity domain in the sense of Koecher(3) while the x is unrestricted. G then contains a subgroup G_∞ of transformations $z \to Az + b$; where b is a real column vector and A belongs to a group of matrices that act transitively on the positivity domain. We now assume that the intersection of Γ with G_∞, which we call Γ_∞, is such that its pure translations (that is, transformations of the form $z \to z + b$) have compact fundamental domain in the x space and that the group formed by the A's that occur in Γ_∞ has compact fundamental domain in the group of the A's occurring in G_∞ which have determinant 1.

The group Γ_∞ then describes a cusp. If we assume that D has only a finite number of cusps all of this type, then we can show that the contribution of each cusp to the formula for $N(r,\Gamma)$ actually is independent of r. This result also holds for Siegel domains of the second kind with the obvious modifications that have to be made in the hypothesis about Γ_∞. Cusps of this form are of course special cases of the general type referred to earlier as pointlike, the above result, however, does not hold for all pointlike cusps. It is, however, probably true that in general a cusp will contribute a polynomial in r of a certain degree which corresponds to the dimension of the cusp in some sense. For the cusps described above the contribution from each cusp to $N(r,\Gamma)$ need not be a rational number, only the total contribution from all

(1) To appear in the *Annals of Mathematics*.

(2) In the sense of Piatetsky-Shapiro [7].

(3) Koecher [4].

cusps has to be rational. In case of the complex domain $|z_1|^2+\ldots+|z_n|^2<1$, however, the contribution from each cusp is rational. In particular, for n odd and >1 it is identically zero. Similarly for irreducible(1) groups acting on a product of unit circles $|z_i|<1, i=1,2,\ldots,n$, the contribution of the cusps (if they are all of the required type) is again zero for n odd and >1. The hyperbolic plane occupies here, as in so many other instances, a somewhat singular position. In general the contribution from a cusp of this type is evaluated as the value of a certain dirichlet-series for some integral value of the argument, and to express the total contribution of the cusps in rational terms is in general a problem in itself; for the case of the so-called Hilbert modular group acting on a product of unit circles, Shimizu(2) has recently given such an expression.

4. Let us suppose that there is an element of G, say g, which is not in Γ, but such that the intersection of $g^{-1}\Gamma g$ and Γ has finite index in both (the index of course then has to be the same), then if $f(z)$ is automorphic under Γ, we have also that $f(gz)$ is automorphic under $\Gamma'=g^{-1}\Gamma g\cap\Gamma$. Taking a set of representatives γ_i of the cosets such that $\Gamma=\sum\Gamma'\gamma_i$, we find that $\sum_i f(g\gamma_i z)$ is again automorphic with regard to Γ. As an operation on f this commutes with our class of integral operators, and the combination of this operation with the integral operators, leads to a more general type of integral operator whose trace we again may try to evaluate in two different ways. This leads to a more general type of trace formula.

Of particular interest in this direction, especially from the point of view of number theory, are the type of operators known as Hecke operators in connection with arithmetic groups Γ. We have again two types of trace formula, the infinite and the finite case, and of course again the great difficulties pertaining to the case of non-compact fundamental domain with finite volume.

I have earlier(3) given as an example of the finite case the formula for the trace of the Hecke operators for the analytic forms of weight k for the classical modular group in the upper complex half plane. It is possible to give analogous formulas for some higher dimensional cases where only the simpler type of cusps described in the last part of the preceding section occurs. For the Hilbert modular groups these formulas are not hard to establish, and are actually in one way simpler than for the classical modular group, particularly if the totally real field contains units of norm -1, since the cusps then bring no contribution to the trace.

It may in passing be remarked that this machinery of course works if the group G is compact. S is then also compact and Γ finite; the theory of Hecke operators, while of course much simpler in this case, is by no means trivial and one can obtain results that are of definite numbertheoretical interest.

Another application of these methods is to the study of dirichlet-series that are formed with the eigenvalues of the Hecke operators as coefficients for an automorphic function which is an eigenfunction for the full ring of Hecke operators. In the case of the classical modular group the properties of these series were derived from the fact that these eigenvalues were

(1) See Selberg [11], p. 163.

(2) To appear in *Annals of Mathematics*.

(3) Selberg [8], p. 85.

essentially also the fourier coefficients of the automorphic function. However, such an interpretation is clearly not always possible, certainly not if the fundamental domain is compact. By suitable combination of the Hecke operators with certain integral operators, however, it is possible to prove that these dirichlet-series possess similar analytic properties (in particular, functional equations), also in a number of cases where the eigenvalues of the function with regard to the Hecke operators cannot be interpreted as fourier coefficients, both in cases where G is noncompact and D compact, and in cases where G is compact.

It is likely that the further study of Hecke operators from these points of view will yield many results of considerable interest for the theory of numbers.

5. The principal open problems pointed out earlier are those connected with noncompactness. A solution of these at least for the arithmetic subgroups of algebraic groups would undoubtedly produce results of great interest for the theory of numbers. Probably of at least equal importance, from the point of view of number theory, is the carrying over of these ideas and methods to groups defined over p-adic number fields and to the so-called adele groups.

REFERENCES

[1]. Borel, A. & Harish-Chandra, Arithmetic subgroups of algebraic groups. *Ann. Math.*, 75 (1962), 485–535.

[2]. Godement, R., *Séminaire Bourbaki*, février, 1957.

[3]. Hirzebruch, F., Automorphe Formen und der Satz von Riemann-Roch. *Symposium internacional de topologia algebraica*. Mexico City 1958, 129–144.

[4]. Koecher, M., Positivitätsbereiche im R^n. *Amer. J. Math.*, 79 (1957), 575–596.

[5]. Maass, H., Über eine neue Art von nichtanalytischen automorphen Funktionen. *Math. Ann.*, 121 (1949), 141–183.

[6]. Piatetsky-Shapiro, I. I., Discrete subgroups of analytic automorphisms of a polycylinder and automorphic forms. *Dokl. Akad. Nauk SSSR.*, 124 (1959), 760–763.

[7]. —— *The Geometry of Classical Domains and the Theory of Automorphic Functions*. Moscow, 1961.

[8]. Selberg, A., Harmonic analysis and discontinuous groups. *J. Ind. Math. Soc.*, 20 (1956), 47–87.

[9]. —— Automorphic functions and integral operators. *Seminars in Analytic Functions* (Princeton, 1957), vol. 2, 152–161.

[10]. —— A new type of zetafunctions connected with quadratic forms. *Report of the Institute in the Theory of Numbers*, Boulder, Colorado, 1959, 207–210.

[11]. —— On discontinuous groups in higher-dimensional symmetric spaces. *Contributions to Function Theory*, 147–164, Bombay, 1960.

[12]. Tamagawa, T., On Selberg's trace formula. *J. Fac. Sci., Univ. Tokyo*, 8 (1960), 363–386.

GÉOMÉTRIE ALGÉBRIQUE

Par JEAN-PIERRE SERRE

1. Introduction

Je voudrais exposer ici quelques uns des développements récents de la géométrie algébrique. Je dois préciser que je prends ce dernier terme au sens qui est devenu le sien depuis quelques années : celui de *théorie des schémas*.

Il n'est pas question de rappeler la définition précise d'un schéma; je renvoie pour cela à la conférence de Grothendieck au congrès précédent. Disons seulement que, alors qu'une variété algébrique se construit à partir d'algèbres de polynômes sur des corps, un *préschéma* se construit à partir d'anneaux commutatifs quelconques (et un schéma est un préschéma séparé)(1). Bien entendu, certains théorèmes exigent des hypothèses de finitude (nous en verrons de nombreux exemples), mais ces hypothèses sont vérifiées par des anneaux que l'ancienne géométrie algébrique ne pouvait traiter qu'indirectement (et incomplètement), ne serait-ce que l'anneau $\mathbf{Z}$ des entiers. (Et pourtant, on peut dire qu'une partie importante de l'arithmétique consiste justement en l'étude des schémas de type fini sur $\mathbf{Z}$ — c'est là, par exemple, le cadre naturel de la théorie des fonctions zêta et L, à la Artin-Weil.) Un autre type d'anneaux qui intervient fréquemment est celui des *anneaux locaux* : lorsqu'on désire étudier un morphisme $f:X\to Y$ au voisinage de $f^{-1}(y)$, avec $y\in Y$, on est amené à remplacer X par un schéma X' sur l'anneau local O_y de y, celui déduit de X par image réciproque au moyen du morphisme $\mathrm{Spec}(O_y)\to Y$; souvent même, pour localiser davantage, on remplace O_y par son complété $\hat{O}_y$.

Je n'essaierai pas de donner un aperçu complet des résultats obtenus ces dernières années en théorie des schémas, et je me limiterai aux deux thèmes suivants : *théorèmes d'existence*, et étude des *schémas sur un anneau local noethérien complet*. Pour le reste, on pourra se reporter aux exposés de Grothendieck au séminaire Bourbaki [7, 10] et au séminaire de l'I.H.E.S. [9], ainsi, bien entendu, qu'aux *Eléments* [8].

2. Théorèmes d'existence

Il s'agit chaque fois de construire un préschéma M de propriétés données; le plus souvent, comme Grothendieck l'a mis en évidence, ces propriétés s'expriment de la façon la plus commode en disant que M *représente* un certain foncteur. Rappelons ce que l'on entend par là :

Soit C une catégorie, soit $\mathcal{E}ns$ la catégorie des ensembles, et soit $F:\mathsf{C}\to\mathcal{E}ns$ un foncteur contravariant. On dit qu'un couple (M,m), où $M\in\mathrm{Ob}(\mathsf{C})$, $m\in F(M)$, *représente* F si, pour tout $T\in\mathrm{Ob}(\mathsf{C})$, l'application de $\mathrm{Hom}(T,M)$ dans $F(T)$ qui associe à $\varphi\in\mathrm{Hom}\,(T,M)$ l'élément $F\varphi(m)$ de $F(T)$ est une

(1) D'après Grothendieck, il faudrait élargir cette définition, de telle sorte que l'on puisse parler de « préschémas au-dessus d'un espace annelé », cet espace pouvant être, par exemple, une variété différentiable ou analytique (comme dans les travaux de Kodaira-Spencer sur les variétés de modules).

bijection; si $h_M : \mathcal{C} \to \mathcal{E}ns$ désigne le foncteur $h_M(T) = \mathrm{Hom}(M, T)$, il revient au même de dire que m définit un isomorphisme de h_M sur F. Lorsqu'un tel couple (M, m) existe, il est unique (à isomorphisme unique près), et l'on dit que F est *représentable* (on dit aussi, par abus de langage, que M représente F). En géométrie algébrique, on prend le plus souvent pour catégorie $\mathcal{C}$ la catégorie $\mathcal{S}ch/_S$ des préschémas au-dessus d'un préschéma de base S (i.e. munis d'un morphisme dans S). On ne dispose pas de critères généraux maniables (1) permettant d'affirmer qu'un foncteur F donné est représentable. Il est en tout cas nécessaire que F transforme limites inductives (dans $\mathcal{S}ch/_S$) en limites projectives (dans $\mathcal{E}ns$); en particulier, F doit être un foncteur *local* : si U parcourt l'ensemble des ouverts d'un S-préschéma T, le préfaisceau des $F(U)$ doit être un *faisceau*.

Exemples de schémas définis par ce procédé :

(i) *Schémas de Grassmann.* On se donne un schéma S, un O_S-Module quasi-cohérent $\mathcal{E}$, et un entier $n \geqslant 0$. Si T est un S-préschéma, notons $\mathcal{E} \times_S T$ l'image réciproque de $\mathcal{E}$ par le morphisme canonique de T dans S, et soit $F_n(T)$ l'ensemble des *quotients du O_T-Module $\mathcal{E} \times_S T$ qui sont localement libres de rang n*; c'est un foncteur contravariant de T (sur la catégorie $\mathcal{S}ch/_S$)). On démontre sans grandes difficultés (cf. Grothendieck [11], exposé 12, pour le cas analytique) que ce foncteur est représentable par un S-préschéma $\mathrm{Grass}_n(\mathcal{E})$. que l'on appelle la *grassmannienne* (d'indice n) de $\mathcal{E}$; si $\mathcal{E}$ est de type fini, $\mathrm{Grass}_n(\mathcal{E})$ est projectif sur S; si de plus $\mathcal{E}$ est localement libre, $\mathrm{Grass}_n(\mathcal{E})$ est simple sur S. Pour $n = 1$, on retrouve le *schéma projectif* $\mathbf{P}(\mathcal{E})$ associé à $\mathcal{E}$ (cf. [8], II–4.1.1); on définit de même les *schémas de drapeaux* de $\mathcal{E}$.

(ii) *Schémas de Hilbert.* Soit S un schéma noethérien, et soit X un schéma projectif sur S; soit $\mathcal{C}$ la catégorie des S-préschémas localement noethériens. Si $T \in \mathrm{Ob}(\mathcal{C})$, soit $X_T = X \times_S T$, et soit $F(T)$ l'ensemble des *sous-préschémas fermés de X_T qui sont plats sur T*; c'est un foncteur contravariant de T. Grothendieck a démontré que ce foncteur est représentable par un schéma $\mathrm{Hilb}_{X/S}$ qui est somme disjointe de S-schémas projectifs $\mathrm{Hilb}^P_{X/S}$, indexés par certains polynômes à coefficients rationnels (cf. [10], exposé 221 — voir aussi [11], exposé 16, pour le cas analytique). Ces schémas jouent un rôle analogue à celui des classiques « coordonnées de Chow » (qui, elles, servent à paramétrer des *cycles*, et non des sous-schémas). Leur existence entraîne facilement celle d'autres schémas importants (schémas de morphismes, d'isomorphismes, etc. — voir [10], *loc. cit.*); elle intervient également dans la construction des schémas de Picard et des schémas de modules (voir ci-dessous).

(iii) *Schémas de Picard.* (Ici, la définition générale du foncteur F est délicate; je me limiterai à un cas particulier relativement simple.) Soit $\pi : X \to S$ un morphisme de préschémas localement noethériens. On suppose que π est projectif, plat, admet une section, et que ses fibres sont géométriquement intègres. Posons

$$P_{X/S} = H^0(S, R^1\pi(O_X^*)).$$

Vu les hypothèses faites, ce groupe s'identifie à $H^1(X, O_X^*)/H^1(S, O_S^*)$. Si T

(1) On trouvera toutefois dans Grothendieck ([11], exposé 11) des critères permettant de déduire la représentabilité d'un foncteur de celle d'autres foncteurs. Voir aussi [10], exposé 195, pour une caractérisation des foncteurs « proreprésentables ».

est un S-préschéma localement noethérien, posons $F(T)=P_{X_T/T}$, avec $X_T=X\times_S T$; c'est un foncteur contravariant de T. Grothendieck a démontré qu'il est représentable par un S-préschéma en groupes abéliens $\mathrm{Pic}_{X/S}$, appelé le *préschéma de Picard de X sur S* (cf. [10], exposés 232–236 — voir aussi [11], exposé 16, pour le cas analytique). En général, $\mathrm{Pic}_{X/S}$ n'est pas de type fini sur S : lorsque S est le spectre d'un corps, et que X est simple sur S, c'est une extension du « groupe de Néron-Severi » de X (considéré comme groupe discret) par la « variété de Picard » de X, au sens classique; toutefois Mumford a démontré qu'il est somme disjointe de schémas quasi-projectifs sur S.

Les hypothèses faites ci-dessus ne sont certainement pas nécessaires pour l'existence du préschéma de Picard; par exemple Murre a construit $\mathrm{Pic}_{X/k}$ pour tout schéma X propre sur un corps k.

(iv) *Schémas de modules* (courbes de genre donné). Si S est un préschéma, une *courbe de genre g sur S* est un S-préschéma X qui est simple, propre, et dont les fibres sont des courbes algébriques de genre g, au sens usuel. Pour S et g donnés, les classes (à isomorphisme près) de telles courbes forment un ensemble $F_g(S)$; c'est un foncteur contravariant de S. Toutefois, pour $g\geqslant 1$, le foncteur F_g *n'est pas* représentable. Il est nécessaire de le modifier en remplaçant la notion de « courbe » par celle de « courbe rigidifiée » (c'est-à-dire, en gros, munie de points d'ordre fini de sa jacobienne); voir la définition précise (dans le cas analytique) dans Grothendieck [11], exposé 7. Mumford a démontré, grâce à un théorème convenable de passage au quotient, que les foncteurs ainsi définis sont représentables, et il en a déduit la construction du *schéma des modules* de courbes de genre g (au sens absolu, i.e. « sur $\mathbf{Z}$ »); voir [14, 15]. Dans cette direction, on disposait déjà de nombreux résultats partiels : construction du schéma des modules sur $\mathbf{Q}$ (Baily [1, 2]), étude détaillée des genres 1 et 2 (Igusa [12, 13]), cas analytique (cf. par exemple Bers [3] et Grothendieck [11], exposés 7 et 17). Dans le cas général, il manque encore une « bonne » compactification des schémas de modules sur $\mathbf{Z}$ analogue à celle donnée pour le genre 1 par Igusa [12], et sur $\mathbf{Q}$ par Baily [2].

3. Schémas sur un anneau local noethérien complet

Dans tout ce qui suit, on désigne par A un anneau local noethérien complet, par $\mathfrak{m}$ l'idéal maximal de A, et par k le corps résiduel $A/\mathfrak{m}$.

Si X est un schéma sur A, on lui associe la fibre X_0 du point $\mathfrak{m}\in\mathrm{Spec}(A)$, autrement dit le schéma $X_0=X\otimes_A A/\mathfrak{m}$. Si X est de type fini sur A, X_0 est un *schéma algébrique sur k*; on dit que X_0 se déduit de X par « réduction modulo $\mathfrak{m}$ ». On essaie, autant que possible, de *ramener l'étude de X à celle de X_0* (muni éventuellement de structures supplémentaires).

Un premier pas dans cette direction consiste à réduire X, non pas seulement modulo $\mathfrak{m}$, mais modulo un idéal primaire pour $\mathfrak{m}$, par exemple une puissance de $\mathfrak{m}$. Les schémas $X_n=X\otimes_A A/\mathfrak{m}^{n+1}$ ainsi obtenus ont même espace topologique sous-jacent que X_0; leurs faisceaux structuraux forment un système projectif. L'espace X_0, muni de ce système projectif, est appelé le *complété formel* de X, et noté $\hat{X}$; c'est un « schéma formel », au sens de [8], I – 10; il représente, en quelque sorte, les « voisinages infinitésimaux » de la fibre X_0; c'est l'intermédiaire le plus naturel entre X_0 et X.

La situation est particulièrement favorable lorsque X est *propre* sur A;

dans ce cas, en effet, on peut dire que la connaissance de $\hat{X}$ *équivaut* à celle de X (cf. [8], III–5). De façon plus précise, on a les résultats suivants :

(i) Soit $\mathcal{E}$ un O_X-Module cohérent, soit $\mathcal{E}_n = \mathcal{E} \times_{\mathcal{E}x_X} X_n = \mathcal{E}/\mathfrak{m}^{n+1}\mathcal{E}$, et soit $\hat{\mathcal{E}}$ le système projectif des $\mathcal{E}_n$. On a alors des isomorphismes canoniques :

$$H^q(X, \mathcal{E}) = \varprojlim H^q(X_n, \mathcal{E}_n) = H^q(\hat{X}, \hat{\mathcal{E}}) \quad (q \geqslant 0).$$

(C'est un cas particulier du « théorème des fonctions holomorphes », cf. Grothendieck [8], III–4.)

(ii) Appelons « Module cohérent sur $\hat{X}$ » un système projectif $\mathcal{F} = (\mathcal{F}_n)$ de O_{X_n}-Modules cohérents tels que $\mathcal{F}_n = \mathcal{F}_m \times_{X_m} X_n$ si $m \geqslant n$, et soit $\hat{\mathsf{C}}$ la catégorie abélienne formée par ces systèmes. Soit d'autre part C la catégorie des O_X-Modules cohérents. Alors le foncteur $\mathcal{E} \to \hat{\mathcal{E}}$ défini dans (i) est une *équivalence* de C avec $\hat{\mathsf{C}}$. En particulier, tout Module cohérent sur $\hat{X}$ est isomorphe au complété formel $\hat{\mathcal{E}}$ d'un O_X-Module cohérent $\mathcal{E}$ (défini à isomorphisme unique près). [C'est le « théorème d'existence en géométrie formelle », cf. [8], III-5 — noter son analogie avec celui qui permet de passer de la géométrie analytique à la géométrie algébrique (sur $\mathbf{C}$).]

Ces résultats ont de nombreuses applications. Citons par exemple :

(iii) (cf. [7], p. 182–11) Supposons que X soit *propre* et *plat* sur A, et que $H^2(X_0, O_{X_0}) = 0$. Alors pour tout O_{X_0}-Module inversible $\mathcal{E}_0$ sur X_0, il existe un O_X-Module inversible $\mathcal{E}$ sur X tel que $\mathcal{E}_0 = \mathcal{E} \times_X X_0$. En particulier, si X_0 est projectif sur k, X est projectif sur A (prendre pour $\mathcal{E}_0$ un Module ample).

[On prolonge $\mathcal{E}_0$ aux X_n de proche en proche; *l'obstruction* pour passer de X_{n-1} à X_n se trouve dans $H^2(X_0, O_{X_0}) \otimes \mathfrak{m}^n/\mathfrak{m}^{n+1}$, qui est nul par hypothèse; on obtient ainsi un Module inversible $\hat{\mathcal{E}}$ sur $\hat{X}$, qui est "algébrique" d'après (ii).]

(iv) (Cf. [7], p. 182–24, ainsi que [9].) Supposons encore que X soit *propre* et *plat* sur A. L'application canonique $\pi_1(X_0) \to \pi_1(X)$ est alors un isomorphisme.

[Il faut démontrer que les catégories formées par les revêtements étales de X_0 et de X sont équivalentes; or un revêtement étale de X_0 est défini par une O_{X_0}-Algèbre cohérente $\mathcal{A}_0$ d'un certain type; on montre qu'on peut lui associer, de façon essentiellement unique, un système projectif $\hat{\mathcal{A}} = (\mathcal{A}_n)$ de O_{X_n}-Algèbres cohérentes $\mathcal{A}_n$ du même type; on applique ensuite (ii) comme ci-dessus.]

Du point de vue où nous nous sommes placés, il est naturel d'essayer de reconstruire (1) le schéma formel $\hat{X}$, ou même le schéma X, à partir de X_0 (en exigeant toujours que $\hat{X}$ soit *plat* sur A). Supposons, pour simplifier, que X_0 soit *simple* sur k, autrement dit, que X_0 soit une « variété algébrique non singulière ». La construction de X_n à partir de X_{n-1} se heurte alors à une *obstruction* qui appartient à $H^2(X_0, \mathcal{T}_0) \otimes \mathfrak{m}^n/\mathfrak{m}^{n+1}$, où $\mathcal{T}_0$ désigne le O_{X_0}-Module des k-dérivations de O_{X_0} (celui qui correspond au fibré tangent à X_0). Si $H^2(X_0, \mathcal{T}_0)$ est nul, cette obstruction est nulle, et l'on peut construire le système des X_n, d'où un schéma formel $\hat{X}$; si en outre X_0 est projectif et si $H^2(X_0, O_{X_0})$ est nul, la méthode employée dans (iii) ci-dessus montre que $\hat{X}$ est le complété formel d'un schéma X qui est projectif et plat sur A; c'est le schéma cherché.

(1) La classification des schémas formels ainsi obtenus conduit à la notion de « schéma formel des modules », cf. Grothendieck [7], p. 182–16, ainsi que [10], exposé 195.

Lorsque k est de caractéristique $p>0$, il est intéressant de prendre pour A un anneau de valuation discrète de caractéristique 0 (par exemple l'anneau $W(k)$ des vecteurs de Witt à coefficients dans k, si k est parfait); on « remonte » ainsi X_0 de la caractéristique p à la caractéristique 0 (toujours en supposant la nullité de $H^2(X_0, \mathcal{J}_0)$ et de $H^2(X_0, O_{X_0})$). En particulier, *toute courbe algébrique se remonte*; de ce résultat, joint à (iv), Grothendieck a déduit la détermination de la partie « première à p » du groupe fondamental d'une courbe en caractéristique p (cf. [7], p. 182–27, ainsi que [9]). En dimension supérieure, il existe par contre des schémas projectifs et simples qui ne se remontent pas en caractéristique 0 (même comme schémas formels); on en trouvera des exemples dans [21] (ces exemples sont de dimension $\geqslant 3$, mais Mumford m'a fait observer qu'on peut en construire d'analogues en dimension 2). Ce genre de question mériterait une étude plus approfondie. Peut-on par exemple définir directement (par des arguments de géométrie différentielle) les obstructions mentionnées ci-dessus ? Elles ont sans doute des relations avec la cohomologie de X_0 à valeurs dans les vecteurs de Witt (cf. [18]). Que peut-on dire du relèvement des morphismes (et en particulier des morphismes « de Frobenius ») ? Le seul cas où l'on sache quelque chose est celui des courbes elliptiques (Deuring [4]).

Les méthodes précédentes font intervenir de façon essentielle des *schémas sur des anneaux artiniens*. D'après Greenberg [6], ces derniers peuvent être eux-mêmes ramenés — dans une certaine mesure — à des *schémas algébriques*. De façon plus précise, supposons que le corps résiduel k soit *parfait*, et que l'anneau local A soit de longueur finie N; si k est de caractéristique 0, choisissons un relèvement de k dans A. On peut alors décrire les éléments de A au moyen de N coordonnées à valeurs dans k, les lois de composition étant polynomiales par rapport à ces coordonnées (c'est clair si k est de caractéristique 0, car A est une k-algèbre — en caractéristique p, on utilise le fait que A est une algèbre sur l'anneau $W(k)$ des vecteurs de Witt); ces lois polynomiales définissent un *schéma en anneaux* Λ sur k, et l'on a $\Lambda_k = A$. Soit maintenant C_k (resp. C_A) la catégorie des k-préschémas (resp. des A-préschémas) de type fini. Si $Y \in \mathrm{Ob}(\mathsf{C}_k)$, soit GY le A-schéma obtenu en munissant Y du faisceau des germes de k-morphismes de Y dans Λ. On définit ainsi un foncteur $G: \mathsf{C}_k \to \mathsf{C}_A$ qui a un *adjoint* $F: \mathsf{C}_A \to \mathsf{C}_k$; c'est cet adjoint que l'on appelle le *foncteur de Greenberg*. Par définition, on a

$$\mathrm{Hom}_A(GY, X) = \mathrm{Hom}_k(Y, FX) \quad (Y \in \mathrm{Ob}(\mathsf{C}_k),\ X \in \mathrm{Ob}(\mathsf{C}_A)),$$

d'où, en prenant $Y = \mathrm{Spec}(k)$, la formule $X_A = (FX)_k$.

Revenons maintenant au cas d'un anneau local noethérien complet A, à corps résiduel parfait k, et soit X un schéma de type fini sur A (ou plus généralement un schéma formel de type fini). Pour tout $n \geqslant 0$, on peut appliquer le foncteur de Greenberg au $A/\mathfrak{m}^{n+1}$-schéma X_n défini plus haut; on associe ainsi à X un *système projectif* $FX = (FX_n)$ *de schémas algébriques sur* k; on a encore $X_A = (FX)_k = \varprojlim (FX_n)_k$. On trouvera dans [6] une liste des propriétés élémentaires du foncteur F. Il conviendrait de la compléter sur plusieurs points. Par exemple, si X est simple sur A, Greenberg a montré que chaque FX_n est un espace fibré sur le précédent, et il en a déterminé le groupe structural; lorsque A est un anneau de valutation discrète, de corps des fractions K, un résultat analogue doit être valable (pour n assez grand) lorsqu'on suppose seulement que $X \otimes_A K$ est simple sur K

(cf. Néron [17], ainsi que Bourbaki, *Alg. comm.*, Chap. III, § 4, $n°5$). Il serait également intéressant (mais sans doute plus difficile) d'« enrichir » le système des FX_n de structures supplémentaires permettant de reconstruire le schéma formel de départ.

Lorsqu'on suppose que X est un *schéma en groupes* sur A, les FX_n forment un *système projectif de groupes algébriques*. C'est ainsi que s'introduisent les *groupes proalgébriques*, limites projectives de groupes algébriques. Dans le cas commutatif, ces groupes forment une catégorie abélienne $\mathfrak{P}$, contenant la sous-catégorie $\mathfrak{P}_i$ des groupes proalgébriques « infinitésimaux » (i.e. limites projectives de groupes réduits à l'élément neutre). La catégorie quotient $\mathfrak{P}/\mathfrak{P}_i$ a été étudiée en détail dans [19] (en supposant k algébriquement clos). Quant à la catégorie $\mathfrak{P}_i$, elle constitue le cadre naturel de la théorie des « groupes formels » commutatifs de Dieudonné (cf. Gabriel [5] et Cartier — non publié).

Le *groupe multiplicatif* G_m fournit un exemple simple de ce qui précède : on lui associe un groupe proalgébrique U, et les points de U rationnels sur k correspondent bijectivement aux unités de A. Lorsque A est un anneau de valuation discrète, de corps des fractions K, et de corps résiduel algébriquement clos, le groupe U ainsi obtenu *détermine les extensions abéliennes de* K : le groupe de Galois de l'extension abélienne maximale de K est isomorphe au groupe fondamental $\pi_1(U)$ du groupe proalgébrique U (cf. [20]).

A cet ordre d'idées se rattachent également des résultats extrêmement intéressants de Néron (on les trouvera résumés dans [16] et [17]). Ici, l'on part d'une variété abélienne C sur le corps des fractions K de A, et l'on cherche un schéma en groupes Γ simple sur A tel que $\Gamma \otimes_A K = C$; de façon plus précise, Néron montre(1) que l'on peut choisir Γ de telle sorte que, pour tout schéma X simple sur A, l'application canonique :

$$\mathrm{Hom}_A(X, \Gamma) \to \mathrm{Hom}_K(X \otimes_A K, C)$$

soit une bijection (en d'autres termes, Γ *représente* le foncteur $\mathrm{Hom}_K(X \otimes_A K, C)$ dans la catégorie des schémas simples sur A). En prenant $X = \mathrm{Spec}(A)$, on voit en particulier que C_K s'identifie à Γ_A; grâce au foncteur de Greenberg, ce dernier groupe s'identifie lui-même au groupe $(F\Gamma)_k$ des points rationnels sur k d'un groupe proalgébrique commutatif $F\Gamma$. Ce résultat a été utilisé par Šafarevič et Ogg dans l'étude de la cohomologie galoisienne de C (le corps k étant algébriquement clos). Ce n'est là qu'un début; d'après Grothendieck, il doit exister un « théorème de dualité » qui englobe ces résultats, ceux de Tate (lorsque k est fini), et ceux de la théorie du corps de classes local (au sens classique — et aussi au sens de [20]).

BIBLIOGRAPHIE

[1]. Baily, W., On the moduli of Jacobian varieties and curves. *Inter. Coll. Function Theory*, 51–62. Bombay, 1960.

[2]. —— On the theory of θ-functions, the moduli of abelian varieties, and the moduli of curves. *Ann. Math.*, 75 (1962), 342–381.

[3]. Bers, L., Uniformization and moduli. *Inter. Coll. Function Theory*, 41–49. Bombay, 1960.

(1) Lorsque C est de dimension 1, Néron démontre un théorème plus précis: il plonge Γ dans un schéma $\bar{\Gamma}$ projectif sur A et régulier, jouissant en outre d'une certaine propriété de minimalité (cf. [16, 17]).

[4]. DEURING, M., Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abh. Math. Sem. Hamburg*, 14 (1941), 197–272.

[5]. GABRIEL, P., Sur les catégories abéliennes localement noethériennes et leurs applications aux algèbres étudiées par Dieudonné. *Séminaire d'algèbre et géométrie*. Collège de France, 1960.

[6]. GREENBERG, M., Schemata over local rings. *Ann. Math.*, 73 (1961), 624–648.

[7]. GROTHENDIECK, A., Géométrie formelle et géométrie algébrique. *Séminaire Bourbaki*, 1958–59, exposé 182.

[8]. —— Eléments de géométrie algébrique (rédigés avec la collaboration de J. DIEUDONNÉ), *Publ. Math.* (*I.H.E.S.*), *Paris*, 1960–61–62–...

[9]. —— *Séminaire de géométrie algébrique* (*I.H.E.S.*), *Paris*, 1960–62.

[10]. —— Technique de descente et théorèmes d'existence en géométrie algébrique. *Séminaire Bourbaki*, 1959–62, exposés 190, 195, 212, 221, 232, 236.

[11]. —— Techniques de construction en géométrie analytique. *Séminaire H. Cartan*, 1960–61, exposés 7 à 17.

[12]. IGUSA, J., Kroneckerian model of fields of elliptic modular functions. *Amer. J. Math.*, 81 (1959), 561–577.

[13]. —— Arithmetic variety of moduli for genus 2. *Ann. Math.*, 72 (1960), 612–649.

[14]. MUMFORD, D., An elementary theorem in geometric invariant theory. *Bull. Amer. Math. Soc.*, 67 (1961), 483–486.

[15]. —— *Geometric invariant theory* (en preparation).

[16]. NÉRON, A., Réduction des variétés abéliennes. *J. reine angew. Math.*, 209 (1962), 29–35.

[17]. —— Modèles p-minimaux des variétés abéliennes. *Séminaire Bourbaki*, 1961–62, exposé 227.

[18]. SERRE, J-P., Sur la topologie des variétés algébriques en caractéristique p. *Symp. de top. alg.*, Mexico, 1956, 24–53.

[19]. —— Groupes proalgébriques. *Publ. Math.* (*I.H.E.S.*), *Paris*, 7, 1960.

[20]. —— Sur les corps locaux à corps résiduel algébriquement clos. *Bull. Soc. Math. France*, 89 (1961), 105–154.

[21]. —— Exemples de variétés projectives en caractéristique p non relevables en caractéristique 0. *Proc. Nat. Acad. Sci. U.S.A.*, 47, (1961), 108–109

GROUPES SIMPLES ET GÉOMÉTRIES ASSOCIÉES

Par J. TITS

§ 1. Introduction

1.1. *Groupes finis simples*

Due à Galois [16], la notion de groupe simple (« non décomposable ») est pratiquement aussi ancienne que la notion de groupe elle-même. Galois avait déjà remarqué que le groupe noté actuellement (suivant J. Dieudonné [12]) $\mathrm{PSL}_2(\mathbf{F}_p)$ est un groupe simple, et qu'il n'existe pas de groupe simple d'ordre non premier strictement inférieur à 60. On peut aussi présumer que la simplicité du groupe alterné $\mathfrak{A}_n (n \geqslant 5)$ était connue de lui, bien que non explicite dans ses écrits.

Après Galois, il faut attendre plusieurs décades pour que soient mises en évidence de nouvelles classes infinies de groupes finis simples, par l'étude — due essentiellement à C. Jordan [18] et L. E. Dickson [9] — de certains groupes linéaires sur les corps finis. Quelques années auparavant, E. Mathieu [23] avait découvert les cinq groupes qui portent son nom et dont la simplicité devait être établie plus tard, par G. A. Miller [25] (1).

Les travaux de Jordan et Dickson allaient recevoir un développement considérable avec les recherches de J. Dieudonné et d'autres concernant les groupes classiques sur un corps quelconque (cf. J. Dieudonné [12]). Mais celles-ci, bien que fondamentales pour la théorie générale des groupes simples, laissaient intacte la liste des groupes finis simples connus, et la nomenclature donnée par Dickson au dernier chapitre de son ouvrage [9], complétée par l'adjonction d'une seule nouvelle classe de groupes simples (groupes de type G_2) découverte par Dickson lui-même [10], pouvait encore servir de référence en 1954, plus de cinquante ans après sa publication.

Par contre, depuis 1955, la théorie des groupes finis simples a repris un nouvel essor, à la suite du mémoire fondamental de C. Chevalley [7], et neuf nouvelles classes infinies de groupes simples ont été trouvées par C. Chevalley, R. Steinberg, M. Suzuki, R. Ree et l'auteur (cf. le n° 4.5 ci-dessous).

Il convient de noter ici que si l'on connait actuellement de nombreuses classes de groupes finis simples, le problème de la détermination de tous ces groupes ne semble pas encore près de sa solution, bien qu'on puisse peut-

---

(1) En fait, dans l'article [23] cité, Mathieu traite essentiellement du groupe noté plus loin M_{12} (cf. Tableau 4) et indique seulement en note l'existence du groupe M_{24}, qui fait l'objet d'un article ultérieur [24]. Les autres groupes dits « de Mathieu » M_{11}, M_{22} et M_{23}, sont des sous-groupes des groupes de permutations M_{12} et M_{24} obtenus en fixant certains points. Pour un traitement plus récent et plus simple de ces divers groupes, cf. E. Witt [56].

être considérer comme un premier pas dans cette voie la démonstration par J. Thompson d'une ancienne conjecture d'après laquelle tout groupe fini simple non cyclique est d'ordre pair.

1.2. *Groupes analytiques simples*

Pour les groupes analytiques, la situation est très différente. Ici, la détermination de tous les groupes simples a été accomplie, très tôt dans le développement de la théorie, par W. Killing [19] et E. Cartan [4] pour le cas complexe et, quelques années plus tard, par E. Cartan [5] pour le cas réel.

1.3. *Groupes algébriques*

La synthèse de ces divers aspects de la théorie des groupes simples — groupes finis simples (au moins la partie « descriptive » de la théorie, dans son état actuel), groupes classiques (disons sur un corps commutatif) et groupes analytiques simples — est, dans une large mesure, réalisée au sein de la théorie des groupes algébriques (linéaires(1)), fondée par L. Maurer à la fin du siècle passé et considérablement développée, au cours de ces dernières années, sous l'impulsion de C. Chevalley [6, 8] et A. Borel [3]. En particulier, il s'avère — sans qu'on puisse actuellement en donner une explication quelque peu rationnelle — que tous les groupes finis simples connus, à l'exception des cinq groupes de Mathieu, des groupes alternés et des groupes cycliques d'ordre premier(2), se retrouvent naturellement par l'étude des groupes algébriques simples(3) définis sur les corps finis. Nous verrons que ceci permet de donner à la nomenclature des groupes finis simples en question une présentation largement unifiée.

La présente conférence a pour objet quelques résultats récents de la théorie des groupes algébriques simples, l'accent étant mis sur les résultats les plus significatifs pour la théorie des groupes finis.

§ 2. Définitions. Généralités

Nous commencerons par rappeler la signification de quelques notions fondamentales, en nous bornant souvent à des indications de caractère heuristique; il ne peut être question, en effet, de reprendre ici les définitions précises de tous les termes utilisés.

---

(1) Il ne sera question, dans cet exposé, que de groupes algébriques affins, c'est-à-dire linéaires (cf. n°s 2.3 et 2.4). Des renseignements et des reférences concernant les groupes algébriques en général peuvent êtré trouv s par exemple dans S. Lang [22].

(2) Seuls les groupes de Mathieu constituent un exception sérieuse. En effet, Z_p est le groupe des points rationnels sur le corps premier $\mathbf{F}_p$ du groupe additif, qui a beaucoup de qualités d'un groupe algébrique simple, bien que, pour des raisons techniques, on ne le considère généralement pas comme tel (ni même comme presque simple : cf. n° 2.2). Quant au groupe alterné $\mathfrak{A}_n$, c'est le dérivé du groupe symétrique $\mathfrak{S}_n$, qui joue un rôle fondamental dans la théorie des groupes algébriques simples en tant que groupe de Weyl des groupes de type $A_{n-1} = \mathrm{PGL}_n$, et qui peut d'ailleurs aussi être considéré comme le groupe projectif $\mathrm{PGL}_n(\mathbf{F}_1)$ sur le « corps à un élément » (cf. J. Tits [44]).

(3) Il s'agit bien entendu de groupes algébriques connexes (cf. n° 2.1). Il est clair qui tout groupe fini (simple ou non) peut être considéré, de façon triviale, comme un groupe algébrique discret.

2.1. *Ensembles algébriques*

RÉFÉRENCES(1). A. Borel [3], C. Chevalley [8], S. Lang [21].

Soit k un corps que, pour éviter certaines complications de nature technique, nous supposerons toujours parfait(2), bien que cette restriction ne soit généralement pas essentielle à la validité des résultats exposés.

Soient A un espace affin défini sur k et $\{x_i\}$ un système de coordonnées affines dans A. Selon un point de vue habituel en géométrie algébrique, il y a lieu de considérer comme points de A, non seulement ses « points rationnels sur k », c'est-à-dire à coordonnées dans k, mais plus généralement tous les points à coordonnées dans un « domaine universel », extension algébriquement close de k qui, pour les besoins de cette conférence, pourra toujours être réduit à la clôture algébrique de k.

Une partie de A est dite *k-fermée* si elle est définie par un système d'équations polynomiales en les x_i, à coefficients dans k; les parties k-fermées de A sont les fermés d'une topologie appelée *k-topologie* (de Zariski) Une partie B de A est *k-connexe* (resp. *k-irréductible*) si, étant donnés deux parties k-fermées disjointes (resp. quelconques) F et F' de A, la relation $B \subset F \cup F'$ implique $B \subset F$ ou $B \subset F'$.

Une application f d'une partie k-fermée F d'un espace affin dans un espace affin est appelée un *k-morphisme* (d'ensemble algébrique) si — pour reprendre un abus de langage classique et imagé — les coordonnées de l'image $f(p)$ d'un point $p \in F$ sont des polynômes à coefficients dans k en les coordonnées de p.

Un *ensemble algébrique affin* — dans la suite, nous dirons simplement « ensemble algébrique », parce que nous n'en considérerons pas d'autre — *défini sur k*, est, en substance, une partie k-fermée d'un espace affin considérée intrinsèquement, c'est-à-dire abstraction faite de l'espace affin ambiant(3). Compte tenu des définitions précédentes, on voit immédiatement ce que sont les parties k-fermées d'un ensemble algébrique défini sur k, la connexité ou l'irréductibilité (sur k) d'un tel ensemble, les morphismes d'un ensemble algébrique dans un autre, deux ensembles algébriques isomorphes (sur k).

Lorsqu'on veut insister sur le corps de définition k d'un ensemble algé-

---

(1) En tête de chaque numéro, nous donnons les références de quelques articles ou livres où on peut trouver des renseignements plus détaillés sur les questions abordées. Nous n'avons cependant, en aucun cas, cherché à établir une bibliographie complète des sujets traités. D'autre part, on notera que les articles cités ne couvrent généralement pas totalement les matières exposées; cela vient de ce que celles-ci appartiennent encore partiellement à la tradition orale.

(2) Rappelons qu'un corps est dit parfait s'il est de caractéristique 0, ou s'il est de caractéristique p et contient la racine p-ième de chacun de ses éléments.

(3) Cette définition heuristique des ensembles algébriques peut être traduite en termes précis : Un ensemble algébrique affin défini sur k est caractérisé par la donnée d'un ensemble E et, pour tout espace affin A défini sur k, d'un ensemble $\mathcal{J}(A)$ d'applications injectives de E dans A, appelées *k-immersions fermées*, et soumises aux conditions

(i) Il existe des espaces A tels que $\mathcal{J}(A)$ ne soit pas vide;

(ii) Etant donnée une immersion fermée $f: E \to A$, pour qu'une injection $\varphi: E \to B$ de E dans un espace affin B soit une immersion fermée, il faut et il suffit que $\varphi(E)$ soit k-fermé et que $f \circ \varphi^{-1}: \varphi(E) \to A$ et $\varphi \circ f^{-1}: f(E) \to B$ soient des k-morphismes d'ensembles algébriques.

Notons encore que la définition donnée ici des ensembles algébriques définis sur k n'est pas valable sur un corps non parfait

brique E, on fait précéder le symbole E de l'indice k, ce qui donne la notation ${}_kE$. Si k' est un surcorps de k, l'ensemble algébrique défini sur k' déduit de ${}_kE$ par extension du corps de base est alors noté ${}_{k'}E$. Enfin, on désigne par E_k l'ensemble des points de ${}_kE$ rationnels sur k.

2.2. *Groupes algébriques*

RÉFÉRENCES. A. Borel [3], C. Chevalley [6, 8], S. Lang [22].

Tous les groupes algébriques dont il sera question ici seront des groupes algébriques *affins*.

Un *groupe algébrique* (affin) *défini sur k* est un ensemble algébrique (affin) G défini sur k, doté d'une loi de groupe telle que l'application $q: G \times G \to G$ définie par $q(x, y) = x \cdot y^{-1}$ soit un k-morphisme. Un sous-groupe de G, au sens de la théorie des groupes algébriques, est un sous-groupe au sens « abstrait » qui est, en outre, k-fermé; un tel sous-groupe est donc, à son tour, un groupe algébrique défini sur k. Si H est un sous-groupe invariant (k-fermé) de G, on peut définir, de façon naturelle, un groupe algébrique quotient G/H, défini sur k, et dont le groupe « abstrait » sous-jacent est le quotient des groupes « abstraits » G et H.

Une application d'un groupe algébrique défini sur k dans un autre est un (k-)*homomorphisme* si c'est à la fois un homomorphisme de groupe « abstrait » et un (k-)morphisme d'ensemble algébrique. On appelle *isogénie* un homomorphisme surjectif de noyau fini, et *isogénie radicielle* une isogénie dont le noyau est réduit à l'élément neutre, c'est-à-dire qui est un isomorphisme de groupes « abstraits ». Les isogénies jouent ici le rôle des isomorphismes locaux de la théorie des groupes de Lie(1); cependant, l'existence d'isogénies radicielles qui ne sont pas des isomorphismes est un phénomène spécifique aux groupes algébriques en caractéristique $p \neq 0$.

Un groupe algébrique (resp. analytique) connexe est dit *semi-simple* s'il ne possède pas de sous-groupe invariant résoluble(2) connexe non trivial, *presque simple* s'il n'est pas commutatif et ne possède aucun sous-groupe invariant propre infini (resp. non discret) et *simple* s'il ne possède aucun sous-groupe invariant propre non trivial(3). Le centre C d'un groupe algébrique (resp. analytique) semi-simple G est fini (resp. discret) et le quotient G/C est un produit direct de groupes simples, éventuellement à une isogénie radicielle près; si G est presque simple, G/C est simple.

---

(1) Plus exactement, la relation entre deux groupes algébriques G et G_1 qui doit être mise en parallèle avec l'isomorphisme local est l'existence d'isogénies d'un même groupe sur G et sur G_1.

(2) Rappelons qu'un groupe G est dit *résoluble* s'il possède une suite de sous-groupes invariants emboîtés $G = G_0 \supset G_1 \supset G_2 \supset \ldots \supset G_n = \{e\}$, telle que les quotients G_i/G_{i+1} soient tous commutatifs. Dans le cas d'un groupe G algébrique, ce n'est pas une restriction de supposer les sous-groupes G_i fermés (pour la topologie de Zariski).

(3) L'expérience (et la théorie de Grothendieck) montre que la notion de groupe simple n'est pas une « bonne notion », et qu'on a le plus souvent avantage à lui substituer celle, plus générale, de groupe presque simple, ou au contraire, de se restreindre à la considération des groupes simples « adjoints » dont il sera question plus loin. On verra d'ailleurs que la différence entre les notions de groupe simple et de groupe simple adjoint est assez subtile (aussi subtile qu'une isogénie radicielle!).

2.3. *Exemples. Groupes classiques*

RÉFÉRENCES. C. Chevalley [8], J. Dieudonné [11].

Le groupe GL_n des matrices inversibles $\mathbf{x}=((x_{ij}))$ d'ordre n, à coefficients dans le domaine universel, a une structure naturelle de groupe algébrique. Pour donner une description de cette structure conforme aux définitions des numéros précédents, il ne suffit pas de considérer les x_{ij} comme des coordonnées dans un espace affin A_0 de dimension n^2; en effet, l'ensemble $GL_n \subset A_0$ n'est pas k-fermé (puisque la relation dét $\mathbf{x} \neq 0$, qui exprime que $\mathbf{x}$ est inversible, n'est pas une équation), et de plus, les coordonnées de $\mathbf{x}^{-1}$ ne sont pas des polynômes en les coordonnées de $\mathbf{x}$. Cependant, on remédie simultanément aux deux défauts en adjoignant aux x_{ij} une (n^2+1)-ième coordonnée $y=1/$dét $\mathbf{x}$, ce qui revient à plonger GL_n dans un espace affin à n^2+1 dimensions. On notera ${}_kGL_n$ le groupe GL_n considéré comme défini sur k, c'est-à-dire le groupe algébrique défini sur k obtenu par le procédé qui vient d'être décrit, en considérant l'espace affin A (et les coordonnées x_{ij} et y) comme défini sur k.

L'importance du groupe ${}_kGK_n$ réside notamment dans le fait qu'il est, pour la théorie des groupes algébriques, « Her All-embracing Majesty »(1) dont parle H. Weyl. On montre en effet que

Tout groupe algébrique (affin) défini sur k est isomorphe à un sous-groupe de ${}_kGL_n$, pour n assez grand.

Le groupe des points de ${}_kGL_n$ rationnels sur k est le groupe $GL_n(k)$ des matrices inversibles à coefficients dans k. On définit de façon évidente les sous-groupes ${}_kSL_n$, ${}_kO_n(f)$ — où f désigne une forme quadratique en k —, ${}_kO_n^+(f)$ et ${}_kSp_n$, de ${}_kGL_n$, dont les groupes de points rationnels sont, avec les notations de J. Dieudonné [11], $SL_n(k)$, $O_n(k,f)$, $O_n^+(k,f)$ et $Sp_n(k)$. En ce qui concerne le groupe unitaire $U_n(k,f)$, où f désigne une forme hermitienne relative à un automorphisme involutif σ de k, on peut voir que c'est le groupe des points rationnels sur le corps k_0 ($\subset k$) des invariants de σ, d'un groupe algébrique défini sur k_0 que nous désignerons par ${}^k_{k_0}U_n(f)$(2).

On notera ${}_kPGL_n$ (resp. ${}^k_{k_0}PU_n(f)$) le quotient du groupe ${}_kGL_n$ (resp. ${}^k_{k_0}U_n(f)$) par son centre, et ${}_kPO_n(f)$, ${}_kPO_n^+(f)$ et ${}_kPSp_n$, les sous-groupes de ${}_kPGL_n$, images canoniques des sous-groupes ${}_kO_n(f)$, ${}_kO_n^+(f)$ et ${}_kSp_n$ de ${}_kGL_n$(3).

---

(1) Plus d'ailleurs que son homologue analytique puisqu'il existe des groupes analytiques qui ne possèdent pas de représentation linéaire (de dimension finie) fidèle.

(2) Ce groupe peut être défini comme suit. Soit $\varkappa$ un élement donné de k n'appartenant pas à k_0, de sorte que tout élément de k peut s'écrire $x+\varkappa y$, avec $x, y \in k_0$. Les conditions pour qu'une matrice « variable » $((x_{ij}+\varkappa\, y_{ij}))$ conserve la forme hermitienne f s'expriment par des relations polynomiales (à coefficients dans k_0) en les x_{ij}, y_{ij}. Ces relations définissent l'ensemble algégrique sous-jacent à ${}^k_{k_0}U(f)$, et la loi de groupe résulte de la loi de multiplication des matrices.

(3) Remarquons que les groupes de points rationnels sur k de ${}_kPO_{2n}(f)$, ${}_kPSp_{2n}$ et ${}^k_{k_0}PU_{2n}(f)$ sont respectivement $PGO_{2n}(k,f)$, $PGSp_{2n}(k)$ et $PGU_{2n}(k,f)$, et non, comme on pourrait le croire, $PO_{2n}(k,f)$, $PSp_{2n}(k)$ et $PU_{2n}(k,f)$; ces derniers ne sont d'ailleurs pas, de façon naturelle, les groupes de points rationnels de groupes algébriques définis sur k. Notons aussi que si k est un corps de caractéristique 2, ${}_kPO_{2n}^+(f)$ (resp ${}_kPSp_{2n}$) n'est pas le quotient de $O_{2n}^+(f)$ (resp. ${}_kSp_{2n}$) par son centre (qui est d'ailleurs réduit à l'élément neutre), mais est l'image de ce groupe par une isogénie radicielle. De même, si la caractéristique de k divise n, ${}_kPSL_n={}_kPGL_n$ n'est pas le quotient de ${}_kSL_n$ par son centre, mais l'image de ce quotient par une isogénie radicielle.

2.4. *Groupes algébriques simples, groupes analytiques simples et groupes « abstraits » simples*

RÉFÉRENCES. C. Chevalley [6, 7], J. Dieudonné [11], H. Freudenthal [13]. Cf. aussi les indications bibliographiques du § 4 ci-dessous.

Il nous faut à présent analyser brièvement les relations existant entre les trois notions de groupes simples en question dans le titre ci-dessus.

Tous les groupes algébriques dont il sera question par la suite seront supposés connexes (pour la topologie de Zariski).

A tout groupe algébrique G, défini sur k, est naturellement associé un groupe « abstrait », le groupe G_k des points de G rationnels sur k. La question se pose de la relation qui peut exister entre la simplicité de G et celle de G_k.

Supposons tout d'abord que k soit le corps $\mathbf{C}$ des nombres complexes (respectivement le corps $\mathbf{R}$ des nombres réels). Alors, G_k a une structure naturelle de groupe analytique complexe (resp. réel). Si le groupe algébrique G est simple, il en est de même du groupe analytique $G_{\mathbf{C}}$ (resp. de la composante connexe de l'élément neutre — au sens topologique usuel — $G^{\circ}_{\mathbf{R}}$ de $G_{\mathbf{R}}$), qui est aussi un groupe simple au sens « abstrait ». Réciproquement, tout groupe analytique complexe (resp. réel) simple est le groupe $G_{\mathbf{C}}$ (resp. $G^{\circ}_{\mathbf{R}}$) associé à un groupe algébrique simple G bien déterminé, défini sur $\mathbf{C}$ (resp. sur $\mathbf{R}$)(1). La classification des groupes analytiques simples est donc essentiellement équivalente à la classification des groupes algébriques simples sur $\mathbf{R}$ et $\mathbf{C}$.

Lorsque k est algébriquement clos, le groupe « abstrait » G_k est simple si le groupe algébrique G l'est(2).

Il n'en est plus du tout ainsi quand k n'est pas algébriquement clos. Un exemple particulièrement frappant de ce qui peut se passer dans ce cas est fourni par le groupe simple $G={}_{\mathbf{Q}}O_3^+(f)$, groupe des rotations à trois variables sur le corps $\mathbf{Q}$ des nombres rationnels, relatif à un forme quadratique f anisotrope (c'est-à-dire ne possédant pas de zéro non trivial dans $\mathbf{Q}$); J. Dieudonné a montré, en effet, que le groupe $G_{\mathbf{Q}}=O_3^+(\mathbf{Q},f)$ possède une suite décroissante de sous-groupes distingués d'indice fini, $G_{\mathbf{Q}}=G^{(0)}\supset G^{(1)}\supset \ldots \supset G^{(i)}\supset \ldots$, telle que $G^{(i)}/G^{(i+1)}$ soit commutatif pour tout $i>0$, et que l'intersection des $G^{(i)}$ soit réduite à l'élément neutre(3). Cependant, la situation n'est pas toujours aussi défavorable et dans beaucoup de « bons » cas, le groupe G_k des points rationnels d'un groupe algébrique simple G est « presque » simple, en ce sens qu'il possède un « gros » sous-groupe invariant simple. Ainsi, lorsque k est un corps fini (mises à part quelques exceptions : v. plus loin), ou encore pour beaucoup de groupes classiques, le groupe dérivé G'_k de G_k est simple, et peut être considéré comme un « gros » sous-groupe de G_k, ceci pour diverses raisons dans lesquelles nous ne pouvons pas

---

(1) Dans le cas complexe, cette dernière assertion reste valable si on remplace « simple » par « semi-simple ». Dans le cas réel, par contre, deux groupes algébriques semi-simples non isomorphes peuvent donner lieu à un même groupe analytique (exemple : ${}_{\mathbf{R}}SL_3$ et ${}_{\mathbf{R}}PGL_3$), et il existe des groupes analytiques semi-simples qui ne proviennent d'aucun groupe algébrique (exemple: le revêtement universel de $SL_2(\mathbf{R})$).

(2) Ce fait n'est pas évident *a priori*; il signifie, en effet, que si un groupe algébrique G possède un sous-groupe « abstrait » invariant propre non trivial, il en possède aussi un qui soit k-fermé.

(3) Pour d'autres exemples de situations analogues, on peut se reporter à l'ouvrage cité de J. Dieudonné.

entrer ici (notons seulement que, dans le cas d'un corps fini, l'indice de G'_k dans G_k est effectivement « très petit » par rapport à l'ordre de G'_k : cf. le n° 4.5 ci-dessous)(⁰).

§ 3. Groupes algébriques sur un corps algébriquement clos

3.1. *Classification des groupes simples*

RÉFÉRENCES. E. Cartan [4], C. Chevalley [8].

Les groupes analytiques complexes simples ont été déterminés par W. Killing et E. Cartan. Ce sont les *groupes classiques*

$\mathbf{C}A_n = \mathbf{C}PGL_{n+1}$, de dimension $n(n+2)$, avec $n=1,2,3,\ldots$
$\mathbf{C}B_n = \mathbf{C}PO_{2n+1}$, de dimension $n(2n+1)$, avec $n=2,3,4,\ldots$(¹),
$\mathbf{C}C_n = \mathbf{C}PSp_{2n}$, de dimension $n(2n+1)$, avec $n=3,4,5,\ldots$(¹),
$\mathbf{C}D_n = \mathbf{C}PO^+_{2n}$, de dimension $n(2n-1)$, avec $n=4,5,6,\ldots$(¹),

et les cinq *groupes exceptionnels*

$$\mathbf{C}G_2,\ \mathbf{C}F_4,\ \mathbf{C}E_6,\ \mathbf{C}E_7,\ \mathbf{C}E_8,$$

de dimensions respectives 14, 52, 78, 133, 248.

C. Chevalley a montré qu'on a, pour les groupes algébriques simples sur un corps algébriquement clos k quelconque, des résultats tout à fait parallèles. De façon précise,

Il existe des groupes simples ${}_kA_n = {}_kPGL_{n+1}$, ${}_kB_n = {}_kPO_{2n+1}, \ldots, {}_kE_8$, *appelés groupes (simples)* adjoints, *qui sont les « analogues sur k» des groupes analytiques complexes simples (en particulier, chacun de ces groupes a la même dimension que le groupe analytique complexe correspondant). Ces groupes sont, à des isogénies radicielles près*(²), *les seuls groupes algébriques simples définis sur k; en outre, les analogues sur k de deux groupes analytiques complexes simples non isomorphes sont aussi non isomorphes (cependant, si k est un corps de caractéristique* 2, *il existe une isogénie radicielle de* ${}_kB_n$ *sur* ${}_kC_n$).

Soulignons ici le caractère inattendu des résultats précédents. On sait que les méthodes de Killing et Cartan sont purement infinitésimales et relèvent en fait de la théorie des algèbres de Lie. Or, en caractéristique $p \neq 0$, contrairement à ce qui se passe en caractéristique 0, l'algèbre de Lie d'un groupe algébrique ne reflète que très imparfaitement la structure de celui-ci (³). D'ailleurs, si la classification des algèbres de Lie simples sur un corps algébriquement clos k de caractéristique $p \neq 0$ n'est pas encore faite, on

(⁰) (Ajouté aux épreuves.) Dans cet ordre d'idées, on a encore le résultat suivant (non publié) : si un groupe algébrique simple ${}_kG$ possède des éléments unipotents rationnels sur k (toujours supposé parfait), ceux-ci engendrent un groupe « abstrait » simple (rappelons qu'un élément d'un groupe algébrique G est dit *unipotent* si la matrice qui lui correspond dans une représentation linéaire rationnelle fidèle de G a toutes ses valeurs propres égales à 1).

(¹) Les groupes $\mathbf{C}B_1 = \mathbf{C}PO_3$, $\mathbf{C}C_1 = \mathbf{C}PSp_2$, $\mathbf{C}C_2 = \mathbf{C}PSp_4$, $\mathbf{C}D_1 = \mathbf{C}PO^+_2$, $\mathbf{C}D_2 = \mathbf{C}PO^+_4$ et $\mathbf{C}D_3 = \mathbf{C}PO^+_6$ sont respectivement isomorphes à $\mathbf{C}A_1$, $\mathbf{C}A_1$, $\mathbf{C}B_2$, $\mathbf{C}^*$ (groupe multiplicatif du corps $\mathbf{C}$), $\mathbf{C}A_1 \times \mathbf{C}A_1$ et $\mathbf{C}A_3$.

(²) Cette restriction qui s'applique seulement en caractéristique $p \neq 0$ est due au fait que nous parlons ici de groupes simples, et illustre ce qui a été dit dans la note (3), p. 200. Les résultats de Chevalley concernent en fait les groupes semi-simples, pour lesquels le parallélisme entre le cas complexe et le cas d'un corps algébriquement clos quelconque est total.

(³) Ceci est à l'origine de la théorie des hyperalgèbres de Lie, de J. Dieudonné [12].

sait cependant qu'elle est très différente de la classification en caractéristique 0, puisque les « analogues sur k » des algèbres simples complexes ne sont pas nécessairement simples, et que, d'autre part, il existe de nouvelles algèbres simples, sans équivalent en caractéristique 0. En schématisant considérablement la réalité, et en faisant abstraction de l'abondance de notions et de propriétés nouvelles (même dans le cas complexe) mises en évidence par la théorie de Chevalley, on peut dire que celle-ci repose sur une transposition au plan global des méthodes infinitésimales de Killing et Cartan, le fait remarquable étant que, contrairement à la théorie infinitésimale, la théorie globale résiste à un changement de caractéristique.

Dans la suite, les seuls groupes simples définis sur des corps algébriquement clos dont il sera question sont les groupes adjoints; le qualificatif « adjoint » sera toujours omis, de sorte que, pour des groupes définis sur un corps algébriquement clos, l'expression « groupe simple » devra toujours être prise dans le sens de « groupe simple adjoint ».

3.2. *Sous-groupes de Borel et sous-groupes paraboliques*

RÉFÉRENCES. A. Borel [3], C. Chevalley [8], J. Tits [48, 52].

Nous parlerons plus loin de la classification des groupes simples sur un corps non algébriquement clos, mais pour être en mesure d'énoncer les résultats que nous avons en vue à leur propos, nous devons au préalable donner quelques indications supplémentaires sur la structure des groupes algébriques connexes définis sur un corps k algébriquement clos.

Soit G un tel groupe. On appelle *sous-groupe de Borel*, tout sous-groupe résoluble connexe maximal de G, et *sous-groupe parabolique* tout sous-groupe contenant au moins un sous-groupe de Borel (1).

Illustrons ces définitions par un exemple. Soit G le groupe ${}_k\mathrm{PGL}_{n+1}$ des projectivités d'un espace projectif S_n, à n dimensions, défini sur k. Rappelons qu'un ensemble $\{a_1, \ldots, a_m\}$ de sous-variétés linéaires de S_n emboîtées, c'est-à-dire telles que $a_1 \subset a_2 \subset \ldots \subset a_m$, est appelé un *drapeau*, et qu'un drapeau est dit *complet* si les dimensions des variétés linéaires qui le composent prennent toutes les valeurs entières de 0 à $n-1$. Alors, il résulte du théorème de Lie-Kolchin que les sous-groupes de Borel de $G = {}_kPGL_{n+1}$ sont les stabilisateurs des drapeaux complets de S_n. De plus, les sous-groupes contenant le stabilisateur d'un drapeau complet sont les stabilisateurs des parties de ce drapeau; par conséquent, les sous-groupes paraboliques de G sont tous les stabilisateurs de drapeaux de S_n.

Les propriétés suivantes, qui se vérifient aisément dans le cas de l'exemple précédent, sont valables pour tout groupe G.

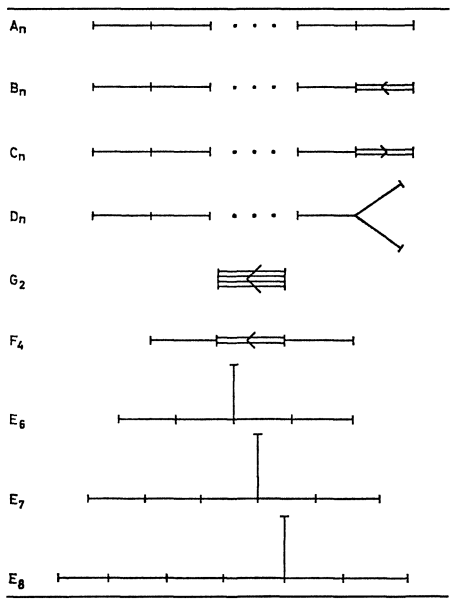
(3.2.1) *Les sous-groupes de Borel de G sont tous conjugués.*

(3.2.2) *Les sous-groupes paraboliques de G sont connexes.*

(3.2.3) *Le treillis des sous-groupes paraboliques contenant un sous-groupe de Borel donné B est isomorphe au treillis des parties d'un ensemble fini. Autrement dit, si $P^{(i)}$ $(i \in I)$ désignent les sous-groupes maximaux de G contenant B, l'ensemble d'indices I est fini et tout sous-groupe contenant B est, de façon unique, une intersection de la forme $P^{(i_1)} \cap P^{(i_2)} \cap \ldots \cap P^{(i_m)}$ $(i_1, i_2, \ldots, i_m \in I)$.*

(1) Conformément à la terminologie fixée au n° 2.2, nous considérons seulement des sous-groupes qui sont fermés pour la topologie de Zariski, mais en l'occurrence les définitions sont indépendantes de cette hypothèse; on montre en effet que tout sous-groupe (« abstrait ») contenant un sous-groupe résoluble connexe maximal est nécessairement fermé.

TABLEAU 1. *Digrammes de groupes simples (formes déployées.)*
(Cf. n°s 3.3 et 4.2.)



Lorsque G est semi-simple, le nombre d'éléments de I est égal au *rang* de G (on appelle ainsi le plus grand entier r tel que G possède un sous-groupe isomorphe au produit direct de r fois le groupe multiplicatif du corps k). Dans l'exemple considéré plus haut ($G = {}_kPGL_{n+1}$), B est le stabilisateur d'un drapeau complet de S_n, les $P^{(i)}$ sont les stabilisateurs des variétés linéaires qui constituent ce drapeau et le rang de G est n.

(3.2.4) *Deux sous-groupes paraboliques contenant un même sous-groupe de Borel B ne sont jamais conjugués; il s'ensuit que les classes de sous-groupes paraboliques conjugués sont en correspondance biunivoque canonique avec les parties de l'ensemble fini I.*

3.3. *Diagrammes*

RÉFERÉNCES. C. Chevalley [8], J. Tits [44, 48, 49], E. Witt [57].

Certaines caractéristiques essentielles d'un groupe algébrique semi-simple G défini sur un corps algébriquement clos k peuvent se résumer par la donnée d'une figure formée de points et de traits éventuellement orientés, appelée le *diagramme de G* (on lui donne aussi fréquemment le nom de « diagramme de Dynkin », bien que son introduction revienne primitivement à H. S. M. Coxeter). Le diagramme d'un groupe simple G dépend seulement du type $A_n, B_n, \ldots$, auquel appartient G et non du corps de base k; les diagrammes des divers types de groupes simples sont donnés dans le Tableau 1. Le diagramme d'un groupe semi-simple G quelconque est une réunion disjointe de diagrammes de groupes simples, G étant alors, à une isogénie près, le produit direct des groupes simples en question.

La définition habituelle du diagramme d'un groupe repose essentiellement sur la notion, assez technique, de « racine » du groupe. Nous en indiquerons une autre, à vrai dire assez artificielle, pratiquement peu maniable

et d'ailleurs incomplète (pour ce qui concerne l'orientation des traits multiples), mais qui offre ici l'avantage de s'exprimer en peu de mots à partir des notions introduites au numéro précédent, et qui nous suffira pour les besoins de cet exposé.

Considérons un groupe G semi-simple(1), et soient B, I et $P^{(i)}$ ($i \in I$) définis comme au n° 3.2. Pour toute partie J de I, désignons par d_J la dimension de l'intersection des groupes $P^{(i)}$, avec $i \notin J$. En particulier, d_ϕ est la dimension de B. Il résulte immédiatement de (3.2.2) et (3.2.3) que si $i, j \in I$, $d_{i,j} - d_\phi \geqslant 2$ (notons au passage que $d_i - d_\phi = 1$ quel que soit i). Cela étant, pour construire le diagramme de G, on représente les éléments de I par des points, appelés *sommets* du diagramme, et on joint les sommets représentant deux éléments quelconques i et j de I par un trait de multiplicité $d_{i,j} - d_\phi - 2$, affecté d'une flèche si cette multiplicité est >1.

Nous ne préciserons pas ici quelle est la signification générale de cette orientation donnée aux traits multiples des diagrammes, mais nous indiquerons, dans le cas particulier d'un groupe G simple de rang 2, à quelle propriété des sous-groupes paraboliques de G elle correspond. Soient $P^{(i)}$ ($i = 1, 2$) les sous-groupes paraboliques maximaux contenant un sous-groupe de Borel donné B de G. Si G est le groupe ${}_kA_2 = {}_kPGL_3$ des projectivités d'un plan projectif, B est le stabilisateur d'un drapeau formé d'un point a_1 et d'une droite a_2 passant par ce point, et $P^{(i)}$ est le stabilisateur de a_i; il résulte alors du principe de dualité de la géométrie projective plane que G possède des automorphismes (extérieurs) conservant B et permutant $P^{(1)}$ et $P^{(2)}$. On n'a rien de semblable lorsque $G = {}_kB_2$ ou ${}_kG_2$; autrement dit, $P^{(1)}$ et $P^{(2)}$ occupent alors des situations asymétriques dans G, et l'orientation du trait constituant le diagramme de G traduit précisément cette asymétrie. Notons encore que celle-ci est quelque peu « atténuée » lorsque k est un corps de caractéristique 2 (resp. 3) et $G = {}_kB_2$(resp. ${}_kG_2$) : dans ce cas, il existe en effet des isogénies radicielles de G sur lui-même qui conservent le sous-groupe de Borel B et permutent $P^{(1)}$ et $P^{(2)}$). De façon générale, on peut dire que, sur un corps de caractéristique 2 (resp. 3), il ne faut pas orienter les traits doubles (resp. quadruples) des diagrammes de groupes semi-simples lorsqu'on considère ceux-ci à une isogénie radicielle près; en caractéristique 2, ceci se traduit encore par l'existence, déjà signalée au n° 3.1, d'isogénies radicielles de ${}_kB_n$ sur ${}_kC_n$, et aussi par l'existence d'isogénies radicielles de ${}_kF_4$ sur lui même qui ne sont pas des automorphismes.

§ 4. Groupes simples sur un corps non algébriquement clos

4.1. *k-formes*

Références. E. Cartan [5], J. P. Serre [35], A. Weil [54].

Soient k un corps (parfait) et K sa clôture algébrique. Etant donné un groupe algébrique ${}_kG$ défini sur k, on en déduit immédiatement, par extension du corps de base, un groupe ${}_KG$ défini sur K. On dit alors que ${}_kG$ est une

(1) L'hypothèse de semi-simplicité ne joue en fait aucun rôle ici; la construction décrite permet donc, en principe, d'associer un diagramme à tout groupe algébrique connexe G, mais il s'avère que ce diagramme n'est autre que celui du groupe semi-simple quotient de G pas son radical (on appelle ainsi le plus grand sous-groupe invariant résoluble connexe de G).

k-forme de ${}_KG$. Par exemple, les groupes orthogonaux réels de dimension donnée sont des **R**-formes, ou *formes réelles*, du groupe orthogonal complexe de même dimension; les groupes orthogonaux réels correspondant à des formes quadratiques d'indices différents n'étant pas isomorphes (sur **R**), ceci montre qu'un même groupe ${}_KG$ peut posséder plusieurs k-formes différentes.

Nous devrons encore faire appel à un autre aspect de la notion de k-forme, qui fait intervenir le groupe de Galois Γ de l'extension K/k. Dorénavant, on identifiera, au point de vue ensembliste, un groupe algébrique G avec l'ensemble de ses points rationnels sur K (ce qui revient à prendre K comme domaine universel). Le groupe ${}_kG$ étant donné, Γ opère sur G de façon évidente : le transformé d'un point g de G par un élément γ de Γ a pour coordonnées (dans une k-immersion affine quelconque de ${}_kG$) les transformées par γ des coordonnées de g. Il est facile de voir que se donner une k-forme ${}_kG$ d'un groupe ${}_KG$ défini sur K, revient à se donner l'action du groupe de Galois Γ sur G. Par exemple, on peut caractériser une forme réelle d'un groupe algébrique complexe ${}_{\mathbf{C}}G$, en donnant l'involution des couples de points complexes conjuguées relatifs à cette forme; c'est d'ailleurs, transposé des algèbres de Lie aux groupes algébriques, le principe de la méthode utilisée par E. Cartan pour la détermination des algèbres de Lie réelles simples. Notons encore que les points de ${}_kG$ rationnels sur k ne sont autres que les points de G invariants par Γ.

Si un groupe ${}_kG$ est semi-simple, il le reste par extension du corps de base. Par contre, un groupe peut être simple sur k sans être « absolument simple », c'est-à-dire ne posséder aucun sous-groupe invariant propre non trivial défini sur k mais en acquérir sur K; un tel groupe n'est donc pas une k-forme de groupe ${}_KG$ simple. Cependant, la théorie des groupes algébriques simples sur un corps non algébriquement clos s'avère essentiellement équivalente à celle des formes de groupes simples, en ce sens que, modulo des problèmes de pure théorie de Galois, il est possible de ramener la classification des groupes algébriques simples sur k à celle de toutes les formes de groupes simples ${}_kG$, définies sur des extensions finies de k.

La classification des k-formes de groupes simples, notamment sur certains corps k particuliers (corps p-adiques, corps de nombres, etc.), a récemment fait l'objet de nombreux travaux dont il ne nous est pas possible de rendre compte ici(1). Nous donnerons seulement, au n° 4.6, quelques indications sur un résultat relatif à la classification des formes de groupes simples sur un corps quelconque. Mais au préalable, nous parlerons plus en détail de deux types particuliers de k-formes, qui jouent un rôle fondamental dans les applications à la théorie des groupes finis. Ces formes sont généralement appelées formes *de Chevalley* (ou encore *normales*) et formes *extérieures types*; nous proposons ici les dénominations, qui nous paraissent plus commodes, de formes *déployées* et *quasi-déployées*.

4.2. *Formes déployées* (*de Chevalley*)

RÉFÉRENCES. C. Chevalley [7], L. E. Dickson [10, 11], R. Ree [26], G. B. Seligman [34], J. Tits [42, 44].

Un groupe semi-simple ${}_kG$, défini sur k, sera dit *déployé* (sur k) si toute classe de sous-groupes paraboliques (définis sur K) conjugués contient au moins un sous-groupe défini sur k. Notons que, dans ce cas, tout sous-

(1) Sur des questions connexes, cf. notamment la conférence de M. Kneser à ce Congrès.

groupe parabolique P contenant un sous-groupe de Borel B défini sur k est lui aussi défini sur k; en effet, le groupe de Galois de K/k conserve B ainsi que chaque classe de sous-groupes paraboliques conjugués, donc aussi P, puisque celui-ci est le seul sous-groupe de sa classe qui contienne B, d'après (3.2.4).

Quel que soit le corps k, tout groupe algébrique simple défini sur la clôture algébrique K de k possède une et (à un isomorphisme près) une seule k-forme déployée.

La k-forme déployée d'un groupe simple ${}_KG$ apparaît, à beaucoup d'égards, comme la k-forme « la plus semblable » au groupe ${}_KG$ lui-même, c'est-à-dire comme l'« analogue sur k » de ce groupe. Ceci est illustré par l'exemple des groupes classiques ${}_KPGL_{n+1}$, ${}_KPSp_{2n}$ et ${}_KPO_n^+$, dont les k-formes déployées sont respectivement ${}_kPGL_{n+1}$, ${}_kPSp_{2n}$ et le groupe ${}_kPO_n^+(f)$ correspondant à une forme quadratique f d'indice maximum.

Nous noterons ${}_kA_n, {}_kB_n, \ldots, {}_kE_8$, les k-formes déployées des groupes ${}_KA_n$, ${}_KB_n, \ldots, {}_KE_8$, et $A_n(k)$, $B_n(k), \ldots, E_8(k)$ les groupes de points rationnels sur k de ces formes.

4.3. *Formes quasi-déployées (extérieures types)*

RÉFÉRENCES. D. Hertzig [17], R. Steinberg [36], J. Tits [45, 46].

Un groupe semi-simple ${}_kG$ sera dit *quasi-déployé* (sur k) s'il possède des sous-groupes de Borel définis sur k. Il résulte immédiatement d'un théorème de S. Lang [1] que *sur un corps fini, tout groupe semi-simple est quasi-déployé*, d'où l'importance des formes quasi-déployées pour la théorie des groupes finis.

Soient ${}_kG$ un groupe semi-simple quasi-déployé, B un sous-groupe de Borel défini sur k — on montre que ces sous-groupes sont tous conjugués par des éléments de G_k — et $P^{(i)}$ les sous-groupes paraboliques maximaux de G contenant B. Le groupe de Galois Γ de K/k conserve B, donc permute entre eux les $P^{(i)}$. Ceux-ci étant représentés par les sommets du diagramme $\mathcal{D}$ de G, on peut dire que Γ « opère sur $\mathcal{D}$ », et il est clair que les éléments de Γ induisent des automorphismes de ce diagramme.

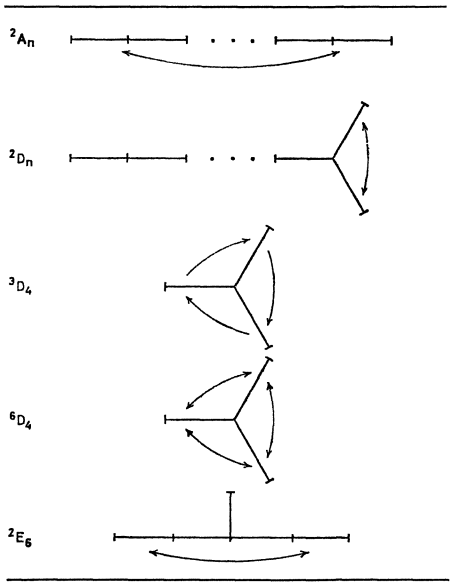
Si Γ opère trivialement sur $\mathcal{D}$, les $P^{(i)}$ sont définis sur k; il en est donc de même, d'après (3.2.3), de tout sous-groupe parabolique contenant B, et le groupe ${}_kG$ est déployé. De façon générale, si Γ_1 désigne le groupe de tous les éléments de Γ qui induisent l'identité sur $\mathcal{D}$, le corps k' des invariants de Γ_1 est la plus petite extension de k telle que ${}_{k'}G$ soit déployé; nous l'appellerons le *corps de déploiement* de ${}_kG$. Notons que ce corps est une extension galoisienne finie de k et que son groupe de Galois $\Delta = \Gamma/\Gamma_1$ opère fidèlement sur $\mathcal{D}$. Le théorème suivant ramène la classification des k-formes quasi-déployées de groupes simples à un problème de pure théorie de Galois.

Soient k un corps quelconque, ${}_KG$ un groupe simple défini sur la clôture algébrique K de k, $\mathcal{D}$ le diagramme de ce groupe, k' un extension galoisienne finie de k, et soit donnée une représentation effective du groupe de Galois Δ de k'/k comme groupe d'automorphismes de $\mathcal{D}$. Alors, ${}_KG$ possède une k-forme quasi-déployée et une seule (à un isomorphisme près) ayant k' comme corps de déploiement et telle que, par rapport à cette forme, Δ opère sur $\mathcal{D}$ de la façon donnée.

Sur le tableau 2 sont représentés, avec des conventions qui s'expliquent d'elles-mêmes, tous les groupes d'automorphismes (non-triviaux) de diagram-

TABLEAU 2. *Groupes d'automorphismes de diagrammes de groupes simples (formes quasi-déployées).*

(Cf. n° 4.3.)



mes de groupes simples. A chacun d'eux correspond, d'après l'énoncé précédent, un type de groupes quasi-déployés. On note qu'une fois connu le type $T = A_n, B_n \ldots$ du diagramme $\mathcal{D}$ considéré et l'ordre δ du groupe Δ, ce groupe et son action sur $\mathcal{D}$ sont bien déterminés, ce qui nous permet de désigner, sans ambiguité, le type de groupes en question par le symbole ${}^\delta T$.

Un groupe quasi-déployé ${}_kG$ de type ${}^\delta T$ donné est bien déterminé (à un isomorphisme près) par la donnée du corps de définition k et du corps de déploiement k'. Nous désignerons ce groupe par ${}^{k'}_{k}T$ et le groupe de ses points rationnels sur k par ${}^\delta T(k, k')$ (la mention explicite de δ nous sera utile, bien qu'explétive en principe, puisque $\delta = [k' : k]$).

Les groupes des types ${}^2A_n$ et ${}^2D_n$ sont des groupes classiques: ${}^{k'}_{k}A_n$ est le groupe ${}^{k'}_{k}PU_{n+1}(f)$ correspondant à une forme hermitienne f d'indice maximum, et ${}^{k'}_{k}D_n$ est le groupe ${}_kPO^+_{2n}(f)$ correspondant à une forme quadratique f d'indice $n-1$ qui devient d'indice maximum n par extension du corps de base de k à k'.

Lorsque k est un corps fini, le groupe de Galois Γ est cyclique, ce qui exclut les formes de type ${}^6D_4$. Notons aussi que, dans ce cas, k' est bien déterminé par la donnée de k et de δ, ce qui nous permettra d'utiliser la notation abrégée ${}^\delta T(k, k') = {}^\delta T(k)$.

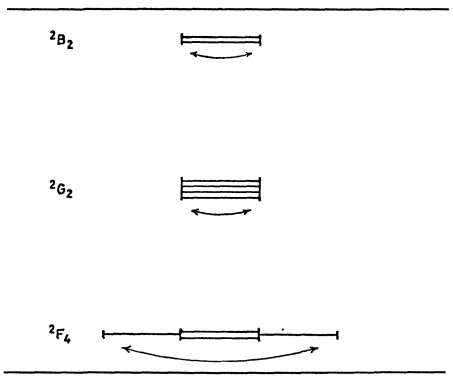
4.4. *Groupes de Ree*

RÉFÉRENCES. R. Ree [27, 28], M. Suzuki [38], J. Tits [50].

Si on ne tient pas compte de l'orientation des traits multiples, la liste des groupes d'automorphismes de diagrammes de groupes simples, donnée par la Tableau 2, doit être complétée par l'adjonction de trois cas, représentés

TABLEAU 3. *Groupes d'automorphismes de diagrammes de groupes simples non orientés (groupes de Ree).*

(Cf. n° 4.4.)



sur le Tableau 3. Nous avons vu au n° 3.3 que si K est un corps algébriquement clos de caractéristique 2 (resp. 3), la symétrie des diagrammes de types B_2 et F_4 (resp. du diagramme G_2) non orienté a une signification pour le groupe simple ${}_KG$ correspondant : elle se traduit par l'existence d'isogénies radicielles exceptionnelles de ce groupe sur lui-même. Ces isogénies étant seulement des automorphismes de groupes « abstraits » et non des automorphismes de groupes algébriques, les diagrammes du tableau 3 ne donnent pas lieu à de nouveaux types de groupes algébriques quasi-déployés. Cependant, R. Ree a montré qu'on peut, sous certaines conditions, leur associer des groupes « abstraits » analogues aux groupes de points rationnels ${}^\delta T(k,k')$ des groupes quasi-déployés.

Ouvrons ici une parenthèse pour fixer un point de terminologie. Par abus de langage, nous dirons qu'un automorphisme du groupe « abstrait » G_k des points rationnels d'un groupe algébrique ${}_kG$ est « algébrique », s'il est induit par un k-automorphisme de ${}_kG$(1), et nous appellerons « groupe d'automorphismes non algébriques » de G_k, tout groupe d'automorphismes de G_k ne contenant aucun automorphisme algébrique à l'exception de l'identité.

Pour introduire les groupes de Ree, revenons tout d'abord aux groupes ${}^\delta T(k,k')$ et voyons comment ceux-ci peuvent être obtenus à partir des groupes $T(k')$ de points rationnels des groupes simples déployés. Soit ${}_kG = {}^{k'}_{k}T$, et soient Γ, Γ_1, Δ et $P^{(i)}$ définis comme au n° 4.3. Le groupe Γ_1 opère trivialement sur le groupe $G_{k'} = T(k')$ des points rationnels du groupe déployé ${}_{k'}G$. Par conséquent, $\Delta = \Gamma/\Gamma_1$ opère sur $G_{k'}$, et le groupe des points de $G_{k'}$ invariants par Δ n'est autre que $G_k = T(k,k')$. On peut voir que l'automorphisme de $G_{k'}$ induit par un élément de Δ distinct de l'élément neutre, n'est pas algébrique. De plus, il est clair que Δ permute les $P^{(i)}_{k'}$ de la façon décrite par la figure ${}^\delta T$ du Tableau 2 (le sens de cette assertion est évident si on se rappelle que les $P^{(i)}$ sont en correspondance canonique avec les sommets du diagramme $\mathcal{D}$ de type T, et que Δ est isomorphe au groupe d'automorphismes de $\mathcal{D}$ représenté par la figure en question).

(1) L'abus de langage consiste essentiellement en ce que le fait, pour un sous-groupe donné, d'être « algébrique » dépend non seulement de la structure de groupe « abstrait » de G_k, mais encore de la façon dont ce groupe est obtenu comme groupe de points rationnels d'un groupe algébrique.

Ces résultats admettent une réciproque. Soient ${}^{\delta}T$ l'un quelconque des « diagrammes avec groupes d'automorphismes » représentés sur le Tableau 2, k' un corps, ${}_{k'}G = {}_{k'}T$ la k'-forme déployée du groupe simple de type T, $P^{(i)}$ les sous-groupes paraboliques maximaux de ${}_{k'}G$ contenant un sous-groupe de Borel donné, défini sur k', $G_{k'}$ et $P^{(i)}_{k'}$ les groupes de points rationnels sur k' de ${}_{k'}G$ et $P^{(i)}$, et Δ un groupe d'automorphismes non algébriques de $G_{k'}$, isomorphe au groupe d'automorphismes du diagramme de type T représenté par la figure ${}^{\delta}T$ du tableau 2, et permutant les $P^{(i)}_{k'}$ de la facon indiquée par cette figure. Alors, le groupe des points fixes de Δ est le groupe ${}^{\delta}T(k,k')$ des points rationnels d'une forme quasi-déployée de G sur un corps $k \subset k'$, qui est le corps des invariants d'un groupe d'automorphismes de k' isomorphe à Δ.

Si l'on cherche à reproduire la construction précédente en partant d'une figure ${}^{2}T$ du tableau 3, on constate que, pour certains corps k', de caractéristique $p=2$ si $T=B_2$ ou F_4, et de caractéristique 3 si $T=G_2$, le groupe $G_{k'}=T(k')$ possède effectivement des automorphismes involutifs π (nécessairement non algébriques) permutant les $P^{(i)}_{k'}$ de la façon indiquée par la figure ${}^{2}T$ en question. Par analogie avec les groupes quasi-déployés, nous appellerons groupe (de Ree) de type ${}^{2}T(*,k')$ les groupes de points fixes d'un tel automorphisme π; la différence essentielle entre le cas des formes quasi-déployées et celui des groupes de Ree est qu'ici, aucun corps k ne joue le rôle de « corps de base ».

Un autre automorphisme involutif π' de $G_{k'}$, jouissant des mêmes propriétés que π, sera dit *équivalent* à π s'il est le transformé de π par un automorphisme intérieur de $G_{k'}$, (auquel cas on peut montrer qu'il existe même un automorphisme intérieur de $G_{k'}$ transformant π en π' et conservant les $P^{(i)}_{k'}$). La classification des automorphismes π en question — donc des groupes de Ree — est donnée par l'énoncé suivant :

Si $T=B_2$ ou F_4 (resp. G_2), les classes d'équivalence d'automorphismes involutifs de $G_{k'}=T(k')$ permutant les $P^{(i)}_{k'}$ de la façon indiquée par la figure ${}^{2}T$ du tableau 3, sont en correspondance biunivoque canonique avec les automorphismes σ de k' tels que

$$(x^{\sigma})^{\sigma}=x^2 \ (resp.\ x^3).$$

Ainsi, à tout automorphisme σ de k' vérifiant cette relation est associé un groupe de type ${}^{2}T(*,k')$, qu'on pourrait par exemple noter ${}^{2}T(\sigma,k')$; nous n'aurons pas ici à faire usage d'une telle notations.

Le corps fini $\mathbf{F}_q$ de $q=p^n$ éléments possède un automorphisme σ tel que $(x^{\sigma})^{\sigma}=x^p$, si et seulement si $n=2m-1$, auquel cas σ ne peut être que la puissance m-ième de l'automorphisme de Frobenius, c'est-à-dire que $x^{\sigma}=x^{p^m}$. L'unicité de cet automorphisme permettra, dans le cas d'un corps k' fini, de parler, sans ambiguité, *du* groupe de Ree ${}^{2}T(*,k')$.

La définition des groupes de Ree donnée ici est, à peu de chose près, celle de R. Ree lui-même. Les groupes ${}^{2}B_2(*,k')$ correspondant à des corps k' finis ont été découverts initialement par M. Suzuki, par une voie très différente.

4.5. *Application: groupes « abstraits » et groupes finis simples*

Références. E. Artin [1], C. Chevalley [7], R. Ree [27, 28], R. Steinberg [36, 37], J. Tits [45, 46, 50].

A de rares exceptions près, tous les groupes « abstraits » dont il a été question dans les trois numéros précédents (4.2, 4.3, 4.4) sont simples, ou « presque ». De façon précise :

Soit G un groupe « abstrait » qui est soit le groupe des points rationnels d'un groupe algébrique simple déployé ou quasi-déployé, soit un groupe de Ree. Alors, le groupe dérivé G' de G est simple, sauf si $G = {}^2A_2(\mathbf{F}_2)$, $A_1(\mathbf{F}_3)$ ou peut-être encore ${}^2F_4(*, F_2)$.(1)

Un cas particulier important est celui où les corps k et k' sont finis. Les groupes en question sont alors des groupes finis simples, et il est remarquable qu'on obtient ainsi la plupart des groupes finis simples connus. En effet,

Les seuls groupes finis simples connus(2) *qui n'appartiennent pas aux types considérés dans l'énoncé précédent sont les groupes alternés $\mathfrak{A}_n$ de degré $n = 7$ ou $\geqslant 9$*(3), *les groupes cycliques Z_p d'ordre premier $p \neq 3$, 5*(4), *et les cinq groupes de Mathieu.*

De plus,

Deux groupes finis simples « de provenance différente » c'est-à-dire représentés, avec les notations introduites plus haut, par des symboles différents, sont « généralement » non isomorphes.

La signification précise de ce dernier énoncé peut être dégagée du Tableau 4, où est reprise la liste détaillée des groupes finis simples connus, avec indication des ordres et de tous les isomorphismes entre groupes d'origine différente. La méthode d'Artin pour la détermination exhaustive de ces isomorphismes repose sur la recherche de toutes les coïncidences entre ordres de groupes finis simples connus, et conduit en outre au résultat suivant :

Dans l'ensemble de tous les groupes finis simples connus, les seules paires de groupes non isomorphes ayant même ordre sont les suivantes:

$$A_2(\mathbf{F}_4)' \quad \textit{et} \quad A_3(\mathbf{F}_2) \simeq \mathfrak{A}_8;$$
$$B_n(\mathbf{F}_q)' \quad \textit{et} \quad C_n(\mathbf{F}_q)', \quad \textit{avec } q \textit{ impair et } n \geqslant 3.$$

4.6. *Un « théorème de Witt » pour les groupes algébriques simples*

RÉFÉRENCES. I. Satake [31], J. Tits [47], E. Witt [55].

Nous avons vu au n° 4.3 comment, étant donné un groupe algébrique semi-simple quasi-déployé ${}_kG$, défini sur un corps k, on peut faire opérer le groupe de Galois Γ de l'extension algébriquement close de k sur le diagramme $\mathcal{D}$ de G. En fait, cette action de Γ sur $\mathcal{D}$ peut être définie pour tout groupe semi-simple ${}_kG$, défini sur k. En effet, d'après (3.2.4), les sommets de $\mathcal{D}$ correspondent biunivoquement, non seulement aux sous-groupes paraboliques maximaux de G contenant un sous-groupe de Borel donné, mais aussi, indépendamment du choix d'un tel sous-groupe, aux classes de sous-groupes paraboliques maximaux conjugués, et ces classes sont évidemment permutées entre elles par le groupe de Galois Γ (qui opère sur G, une fois donnée la forme ${}_kG$ envisagée : cf. le n° 4.1).

Considérons d'autre part les sous-groupes paraboliques de G définis sur k (par rapport à la forme ${}_kG$ en question) minimaux; on montre que ces sous-groupes sont tous conjugués entre eux par des éléments de G_k. D'après

(1) La structure du quotient G/G' est bien connue pour les divers groupes G en question, à l'exception de ${}^2G_2(*, k')$ (cf. les références citées en tête de ce n°).

(2) Selon les informations dont dispose l'auteur en écrivant ces lignes.

(3) $\mathfrak{A}_3 \simeq A_1(\mathbf{F}_2)'$, $\mathfrak{A}_4$ n'est pas simple, $\mathfrak{A}_5 \simeq A_1(\mathbf{F}_4)'$, $\mathfrak{A}_6 \simeq A_1(\mathbf{F}_9)'$, $\mathfrak{A}_8 \simeq A_3(\mathbf{F}_2)$.

(4) $Z_3 \simeq A_1(\mathbf{F}_2)'$, $Z_5 \simeq {}^2B_2(*, \mathbf{F}_2)'$.

TABLEAU 4. *Groupes finis simples.*

(Cf. n° 4.5.)

NOTATIONS:

G et d désignent respectivement un groupe fini et un entier tels que, sauf dans les cas exceptionnels dont la liste est donnée plus loin, G possède un sous-groupe invariant simple G_1 d'indice $(G : G_1) = d$; le sous-groupe G_1 est toujours le groupe dérivé G' de G, sauf quand $G = Z_p$; les groupes G_1 correspondant aux divers groupes G sont tous les groupes finis simples connus; les groupes G_1 correspondant à deux groupes G de provenances différentes, c'est-à-dire représentés par des symboles différents, sont non isomorphes sauf dans les cas énumérés plus bas;

m, n, r = des entiers strictement positifs:

$q = p^r$ = une puissance d'un nombre premier;

(a, b) = *p. g. c. d.* des nombres entiers a et b;

$\mathfrak{S}_n$ (resp. $\mathfrak{A}_n$) = groupe symétrique (resp. alterné) de degré n;

M_n = groupe de Mathieu opérant sur n points (et non sur un nombre de points moindre);

les corps finis qui interviennent dans les symboles $A_n(k)$, ${}^2A_n(k)$, etc. sont désignés par leur ordre (par ex., $A_n(q) = A_n(\mathbf{F}_q)$ etc.); pour la signification des symboles eux-mêmes, cf. n°s 4.2, 4.3 et 4.4.

G	*Ordre de G*	d
$A_n(q)$	$q^{n(n+1)/2}\prod_{i+1}^{n}(q^{i+1}-1)$	$(n+1,\ q-1)$
${}^2A_n(q)$ $(n>1)$	$q^{n(n+1)/2}\prod_{i=1}^{n}(q^{i+1}-(-1)^{i+1})$	$(n+1,\ q+1)$
$B_n(q)$ $(n>1)$	$q^{n^2}\prod_{i=1}^{n}(q^{2i}-1)$	$(2,\ q-1)$
${}^2B_2(*,\ q=2^{2m-1})$	$q^2(q^2+1)(q-1)$	1
$C_n(q)$ $(n>2)$	$q^{n^2}\prod_{i=1}^{n}(q^{2i}-1)$	$(2,\ q-1)$
$D_n(q)$ $(n>3)$	$q^{n(n-1)}(q^n-1)\prod_{i=1}^{n-1}(q^{2i}-1)$	$(4,\ q^n-1)$
${}^2D_n(q)$ $(n>3)$	$q^{n(n-1)}(q^n+1)\prod_{i=1}^{n-1}(q^{2i}-1)$	$(4,\ q^n+1)$
${}^3D_4(q)$	$q^{12}(q^8+q^4+1)(q^6-1)(q^2-1)$	1
$G_2(q)$	$q^6(q^6-1)(q^2-1)$	1
${}^2G_2(*,\ q=3^{2m-1})$	$q^3(q^3+1)(q-1)$	1
$F_4(q)$	$q^{24}(q^{12}-1)(q^8-1)(q^6-1)(q^2-1)$	1
${}^2F_4(*,\ q=2^{2m-1})$	$q^{12}(q^6+1)(q^4-1)(q^3+1)(q-1)$	1

Tableau 4 *(suite)*

G	*Ordre de G*	d
$E_6(q)$	$q^{36}(q^{12}-1)(q^9-1)(q^8-1)(q^6-1)(q^5-1)(q^2-1)$	$(3, q-1)$
${}^2E_6(q)$	$q^{36}(q^{12}-1)(q^9+1)(q^8-1)(q^6-1)(q^5+1)(q^2-1)$	$(3, q+1)$
$E_7(q)$	$q^{63}(q^{18}-1)(q^{14}-1)(q^{12}-1)(q^{10}-1)(q^8-1)(q^6-1)(q^2-1)$	$(2, q-1)$
$E_8(q)$	$q^{120}(q^{30}-1)(q^{24}-1)(q^{20}-1)(q^{18}-1)(q^{14}-1)(q^{12}-1)(q^8-1)(q^2-1)$	1
$\mathfrak{S}_n$ $(n>2)$ (1)	$n!$	2
Z_p	p	1
M_{11}	$7920 = 2^4.3^2.5.11$	
M_{12}	$95040 = 2^6.3^3.5.11$	1
M_{22}	$443520 = 2^7.3^2.5.7.11$	1
M_{23}	$10200960 = 2^7.3^2.5.7.11.23$	1
M_{24}	$244823040 = 2^{10}.3^3.5.7.11.23$	1

EXCEPTIONS

G	*Nature de l'exception*
$\mathfrak{S}_4 \cong A_1(3)$	$G' \cong \mathfrak{A}_4$ n'est pas simple
$A_1(2)$ $(\cong \mathfrak{S}_3)$	$G' \cong Z_3$ est simple mais $(G : G') = 2 \neq d$
${}^2A_2(2)$	G' est résoluble d'ordre 72
$B_2(2)$ $(\cong \mathfrak{S}_6)$	$G' \cong \mathfrak{A}_6$ est simple mais $(G : G') = 2 \neq d$
${}^2B_2(*, 2)$	G est isomorphe au groupe des affinités $(x' = ax+b)$ du corps $\mathbf{F}_5$, donc $G' = Z_5$ est simple, mais $(G : G') = 4 \neq d$
$G_2(2)$	$G' \cong {}^2A_2(3)$ est simple, mais $(G : G') = 2 \neq d$
${}^2G_2(*, 3)$	$G' \cong A_1(8)$ est simple, mais $(G, G') = 3 \neq d$ (2)

(1) Le groupe réduit à l'élément neutre n'est généralement pas considéré comme un groupe simple.

(2) Cet isomorphisme a été signalé à l'auteur par M. Suzuki.

Tableau 4 *(suite)*

ISOMORPHISMES ENTRE GROUPES SIMPLES DE PROVENANCES DIFFÉRENTES

(Les isomorphismes relatifs aux cas exceptionnels énumérés plus haut ne sont pas repris ici.)

$A_1(2)' \cong \mathfrak{S}_3' \cong Z_3$	$A_1(9)' \cong \mathfrak{S}_6'$
$A_1(4) \cong A_1(5)' \cong \mathfrak{S}_5'$	$A_3(2) \cong \mathfrak{S}_8'$
$A_1(7)' \cong A_2(2)$	${}^2A_3(2) \cong B_2(3)'$

REMARQUE

Les types B_1, C_1, C_2, D_1, D_2, D_3, ${}^2D_2$ et ${}^2D_3$ ont été écartés *a priori* de la liste des groupes G considérés. Rappelons, pour mémoire, que

$C_1 \cong B_1 \cong A_1$	$C_2 \cong B_2$	$D_1(q) \cong Z_{q-1}$
$D_2 \cong A_1 \times A_1$	$D_3 \cong A_3$	${}^2D_2(q) \cong A_1(q^2)$
	${}^2D_3 \cong {}^2A_3$	

(3.2.4), les classes de sous-groupes paraboliques de G conjugués correspondent biunivoquement et canoniquement aux ensembles de sommets du diagramme $\mathcal{D}$ de G; en particulier, à la classe qui renferme les sous-groupes paraboliques définis sur k minimaux correspond un certain ensemble de sommets de $\mathcal{D}$, appelés sommets *distingués* pour la forme ${}_kG$ considérée. Notons que, comme on le voit aisément, une classe quelconque de sous-groupes paraboliques conjugués contient des sous-groupes définis sur k si et seulement si l'ensemble de sommets de $\mathcal{D}$ qui lui correspond est invariant par Γ et contenu dans l'ensemble des sommets distingués.

Les groupes quasi-déployés sont caractérisés par le fait que tous les sommets de $\mathcal{D}$ sont distingués; cela résulte immédiatement des définitions. A l'autre extrême, on trouve les groupes ${}_kG$ tels que l'ensemble des sommets distingués soit vide, c'est-à-dire les groupes ne possédant aucun sous-groupe parabolique propre défini sur k; ils sont dits *anisotropes*. Lorsque k est le corps des nombres réels, les groupes semi-simples anisotropes ne sont autres que les groupes semi-simples compacts.

Dans le cas des groupes orthogonaux et unitaires, les notions précédentes — action de Γ sur $\mathcal{D}$ et sommets distingués — sont en relation directe avec l'indice de la forme quadratique ou hermitienne fondamentale. Par exemple, si f est une forme quadratique d'indice i dans un espace vectoriel sur k de dimension impaire $n=2m+1$, les sommets distingués du diagramme B_n pour le groupe ${}_kG={}_kPO_m(f)$ sont les i premiers sommets de ce diagramme, en commençant par l'extrémité opposée au trait double (c'est-à-dire en commençant par la gauche, sur le diagramme B_n du Tableau 1).

On connaît le théorème de Witt qui permet de caractériser une forme quadratique (en caractéristique $\neq 2$) par son indice et sa « partie anisotrope ». On a, pour les groupes algébriques simples sur un corps parfait k, le théorème analogue suivant :

Un groupe simple ${}_kG$ peut être caractérisé par la donnée de l'action de Γ sur le diagramme $\mathcal{D}$ de G, de l'ensemble des sommets de $\mathcal{D}$ distingués pour ${}_kG$, et d'un certain groupe semi-simple anisotrope sur k, la « partie anisotrope de ${}_kG$ », dont le diagramme s'obtient en retirant de $\mathcal{D}$ les sommets distingués pour ${}_kG$ et les traits qui y aboutissent.

§ 5. Interprétations géométriques

5.1. *Géométrie d'incidence associée à un groupe*

RÉFÉRENCES. H. Freudenthal [14, 15], B. A. Rosenfeld [29, 30], J. Tits [39–46, 49, 50, 53].

La signification géométrique des groupes classiques est bien connue. Au cours de ces dernières années, on a beaucoup étudié, de ce même point de vue, les groupes exceptionnels et d'autres groupes introduits par l'étude des groupes algébriques simples, notamment les groupes dont il a été question au § 4. Il ne nous est pas possible de parler ici en détail de toutes les interprétations géométriques qui ont été proposées; cependant beaucoup d'entre elles se rattachent, d'une façon ou d'une autre, aux considérations générales suivantes.

Soit ${}_kG$ un groupe algébrique semi-simple auquel sont associés, d'après les n°s 3.3 et 4.6, un diagramme $\mathcal{D}$, un certain groupe Δ d'automorphismes de $\mathcal{D}$ (isomorphe au quotient du groupe de Galois Γ de l'extension algébriquement close de k par le sous-groupe des éléments de Γ qui induisent l'identité sur $\mathcal{D}$) et un ensemble de « sommets distingués » de $\mathcal{D}$. Soit G_k le groupe des points rationnels sur k de ${}_kG$. Par abus de langage, nous appellerons *sous-groupes paraboliques* de G_k les groupes de points rationnels des sous-groupes paraboliques de ${}_kG$ définis sur k. On montre que les groupes de points rationnels de deux sous-groupes paraboliques de ${}_kG$ définis sur k et distincts (resp. conjugués) sont eux-mêmes distincts (resp. sont conjugués dans G_k). Il s'ensuit, tenant compte de ce qu'on a vu au n° 4.6, que les classes de sous-groupes paraboliques de G_k conjugués correspondent biunivoquement aux ensembles de sommets distingués de $\mathcal{D}$ invariants par Δ; en particulier, les classes de sous-groupes paraboliques maximaux conjugués correspondent aux orbites de Δ dans l'ensemble de tous les sommets distingués.

Tout ceci s'étend immédiatement aux groupes de Ree. Un tel groupe, soit G_*, étant défini comme le groupe des points fixes d'un groupe $G_{k'}$ pour un certain groupe — qu'on notera aussi Δ — d'automorphismes de celui-ci, on peut appeler (toujours par abus de langage) « sous-groupes paraboliques » de G_*, les groupes de points invariants par Δ des sous-groupes paraboliques de G_k, invariants par Δ. Les assertions de l'alinéa précédent relatives aux classes de sous-groupes paraboliques conjugués restent valables ici si on convient, comme il est naturel par analogie avec les groupes quasi-déployés, d'appeler « distingués » tous les sommets du diagramme $\mathcal{D}$ de G.

Dorénavant, G désignera toujours un groupe « abstrait » qui pourra être indifféremment

(5.1.1) le groupe des points rationnels sur un corps k d'un groupe algébrique semi-simple non anisotrope(1) défini sur k, ou bien

(5.1.2) un groupe de Ree.

(1) Les résultats restent valables pour les groupes anisotropes mais deviennent triviaux dans ce cas.

Soient B un sous-groupe parabolique minimal de G, $P^{(i)}$ les sous-groupes paraboliques maximaux contenant B et $E^{(i)}=G/P^{(i)}$ l'espace homogène des classes latérales à droite de $P^{(i)}$ dans G. Un point de $E^{(i)}$ et un point de $E^{(j)}$ seront dits incidents si les classes latérales qu'ils représentent ont une intersection non vide. Les ensemble $E^{(i)}$ et la relation d'*incidence* ainsi définie constituent par définition la *géométrie associée* à G. De façon générale, une *géométrie* (d'incidence) de *rang* r est une entité $\{E^{(i)}; I\}$ constituée par r ensembles $E^{(i)}$ et une « relation d'incidence » I, correspondance symétrique définie sur l'ensemble somme des $E^{(i)}$, et dont la restriction à chaque $E^{(i)}$ est l'identité (i.e., si $x,y \in E^{(i)}$, x et y sont incidents si et seulement si $x=y$). D'après les considérations du début de ce paragraphe, on voit que le rang de la géométrie associée à G est — avec les notations utilisées plus haut — le nombre des orbites du groupe Δ dans l'ensemble des sommets distingués du diagramme $\mathcal{D}$. Remarquons encore que le groupe G opère (par translations à droite) sur la géométrie associée.

A titre d'exemple, supposons que G soit le groupe $PGL_n(k)$ des projectivités (définies sur k) d'un espace projectif S_n à n dimensions défini sur k. Alors, B est le groupe de stabilité d'un drapeau complet $\{a_1, a_2, \ldots, a_n\}$ $(a_1 \subset a_2 \subset \ldots \subset a_n)$ défini sur k, $P^{(i)}$ est le groupe de stabilité de a_i et la géométrie associée à G est la « géométrie projective » dont les ensembles constitutifs $E^{(i)}$ sont l'ensemble des points, l'ensemble des droites, ..., l'ensemble des hyperplans de S_n définis sur k, la relation d'incidence étant la relation d'incidence usuelle (inclusion symétrique) entre sous-variétés linéaires de S_n.

On peut formuler des théorèmes généraux concernant les géométries associées aux groupes des types (5.1.1) et (5.1.2), théorèmes qui traduisent des propriétés structurelles fondamentales communes à tous ces groupes (notamment l'existence pour ceux-ci d'une « décomposition de Bruhat »). Nous nous bornerons ici à indiquer quelques résultats relatifs aux cas très particuliers des géométries de rang 1 et de rang 2.

5.2. *Géométries de rang 1*

RÉFÉRENCES. M. Suzuki [38], J. Tits [50, 51].

La structure d'incidence d'une géométrie de rang 1 est triviale, puisque la relation d'incidence I d'une telle géométrie est nécessairement réduite à l'identité. Cependant, ce cas donne lieu à un résultat significatif, à savoir le fait que

Si la géométrie associée à un groupe G de type (5.1.1) *ou* (5.1.2) *est une géométrie $\{E; I\}$ de rang* 1, *le groupe G est doublement transitif sur E.*

Ceci s'applique en particulier aux groupes de Ree des types ${}^2B_2$ et ${}^2G_2$, qui ont donc des représentations comme groupes doublement transitifs de permutations.

Dans le cas d'un groupe de type ${}^2B_2$, l'ensemble sur lequel le groupe est doublement transitif possède une réalisation « concrète » particulièrement simple. Soient k un corps et $S_n(k)$ l'espace projectif à n dimensions sur k, au sens de la géométrie projective ordinaire, c'est-à-dire l'espace des points rationnels sur k de l'espace projectif S_n considéré par exemple au n° 5.1. Un ensemble non vide $\mathfrak{O}$ de points de $S_n(k)$ est appelé un *ovoïde* si, pour tout point $p \in \mathfrak{O}$, il existe un plan ϖ_p — le « plan tangent à $\mathfrak{O}$ en p » — tel que toute droite passant par p ne rencontre pas $\mathfrak{O}$ en dehors de p ou le rencontre

en un et un seul autre point selon qu'elle appartient ou non ϖ_p. Si k est un corps fini et $n=3$, il résulte d'un théorème de B. Segre [33] que cette notion coïncide avec la notion d'*ovaloïde* introduite par lui. Un ovoïde $\mathfrak{O}$ est *doublement homogène* si le groupe de ses *automorphismes*, c'est-à-dire des projectivités qui le conservent, est doublement transitif sur $\mathfrak{O}$. Enfin.

Tout groupe de type ${}^2B_2(*,k)$ $(k \neq \mathbf{F}_2)$ *est le groupe des automorphismes d'un ovoïde doublement homogène dans l'espace projectif à* 3 *dimensions* $S_3(k)$ *sur* k.

Dans le cas particulier d'un corps k fini, on peut préciser le résultat précédent et en déduire une caractérisation simple des groupes de Suzuki :

Pour que l'espace projectif à 3 *dimensions* $S_3(\mathbf{F}_q)$ *sur le corps fini à* q *éléments contienne des ovoïdes doublement homogènes qui ne soient pas des quadriques ovales, il faut et il suffit que* $q=2^{2m+1}$ $(m>1)$ (1), *auquel cas tous les ovoïdes en question sont projectivement équivalents et ont pour groupe d'automorphismes le groupe de Suzuki* ${}^2B_2(*,\mathbf{F}_q)$.

5.3. *Géométries de rang 2. Géométries polygonales*

RÉFÉRENCES. H. Freudenthal [14], G. J. Schellekens [32], J. Tits [39,46, 49, 50, 53].

Soit S un ensemble dans lequel sont distinguées certaines parties appelées *droites*. Donnons le nom de *polygone à n cotés* à toute « figure » formée par n points $s_1, s_3, \ldots, s_{2n-1}$, tous distincts, les *sommets* du polygone, et par n droites $c_2, c_4, \ldots, c_{2n}$, toutes distinctes, les *côtés* du polygone, tels que $s_i \in c_{i\pm 1}$ et $s_1 \in c_{2n}$. Les droites données définissent dans S une *structure polygonale d'ordre n*, ou *n-gonale*, si les conditions suivantes sont remplies :

S ne contient pas de polygone à $m<n$ côtés;

étant donnés arbitrairement deux points (resp. deux droites; resp. un point et une droite), il existe au moins un polygone à n côtés dont ils sont des sommets (resp. des côtés; resp. un sommet et un côté).

Par exemple, une structure triangulaire est une structure de plan projectif.

La géométrie de rang 2 $\{S, C; I\}$ consituée par un ensemble S et l'ensemble C des droites d'une structure n-gonale dans S, avec la relation d'incidence évidente ($s \in S$ et $c \in C$ sont incidents si $s \in c$), est dite une *géométrie polygonale d'ordre n*, ou *n-gonale*. Alors,

Si le géométrie associée à un groupe G de type (5.1.1) *ou* (5.1.2) *est une géométrie de rang* 2, *elle est polygonale d'ordre* $n=2$, 3, 4, 6 *ou* 8. *De plus, le groupe G est transitif sur les polygones à n côtés de la géométrie, et le sous-groupe des éléments de G conservant un tel polygone induit le groupe complet des automorphismes de celui-ci* (*groupe diédrique d'ordre* $2n$).

Signalons encore, à titre d'exemples, que

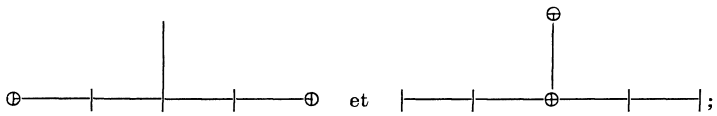
pour les groupes $G_2(k)$, ${}^3D_4(k,k')$ et ${}^6D_4(k,k')$, $n=6$;

les groupes de Ree de type ${}^2F_4$ sont les seuls groupes pour lesquels $n=8$;

si le groupe G dont il est question dans l'énoncé précédent est le groupe des points rationnels d'un groupe (nécessairement non quasi-déployé) de type E_6, sur un corps k de caractéristique $\neq 2$ et 3 (2), l'ensemble des sommets distingués correspondant ne peut être que l'un des deux ensembles représentés par les figures suivantes :

(1) On sait que si q est impair, tout ovoïde dans $S_3(\mathbf{F}_q)$ est une quadrique ovale (cf. A. Barlotti [2]).

(2) Cette restriction n'est peut-être pas nécessaire.



dans le premier cas, $n=3$, et la structure triangulaire associée à G est un plan projectif d'octaves à division sur k; dans le second cas, $n=6$.

Signalons pour terminer que les théorèmes généraux dont il a été question à la fin du n° 5.1, font intervenir des « géométries polyédriques », dont les géométries polygonales considérées ici sont les cas particuliers pour le rang 2.

BIBLIOGRAPHIE

[1]. Artin, E., The orders of the classical simple groups. *Comm. Pure Appl. Math.*, 8 (1955), 455–472.

[2]. Barlotti, A., Un'estensione del teorema di Segre-Kustaanheimo. *Boll. Un. Mat. Ital.* (3), 10 (1955), 498–506.

[3]. Borel, A., Groupes linéaires algébriques. *Ann. Math.*, 64 (1956), 20–82.

[4]. Cartan, E., *Sur la structure des groupes de transformations finis et continus.* Thèse, Paris, 1894; *Œuvres complètes*, 1ère partie, vol. 1, 137–286. Gauthier-Villars, Paris, 1952.

[5]. —— Les groupes réels simples, finis et continus. *Ann. Ec. Norm.*, 31 (1914), 263–355; *Oeuvres complètes*, 1ère partie, vol. 1, 399–491.

[6]. Chevalley, C., Théorie des groupes de Lie. II. (Groupes algébriques). *Actualités Sci. Ind.*, n° 1152, Hermann, Paris, 1951.

[7]. —— Sur certains groupes simples. *Tôhoku Math. J.*, (2), 7 (1955), 14–66.

[8]. —— Classification des groupes de Lie algébriques. *Séminaire C. Chevalley*, (1956–58), Secrétariat mathématique, Paris.

[9]. Dickson, L. E., *Linear Groups with an Exposition of the Galois Field Theory.* B. G. Teubner, Leipzig, 1901.

[10]. —— Theory of linear groups in an arbitrary field. *Trans. Amer. Mat. Soc.*, 2 (1901), 363–394. A new system of simple groups. *Math. Ann.*, 60 (1905), 137–150.

[11]. Dieudonné, J., *La géométrie des groupes classiques.* Ergebnisse d. Math., N.F. 5, Springer. Berlin–Göttingen–Heidelberg, 1955.

[12]. —— Groupes de Lie et hyperalgèbres de Lie sur un corps de caractéristique $p>0$ (alternativement: même titre en anglais). I. *Comm. Math. Helv.*, 28 (1954), 87–118; II. *Amer. J. Math.*, 77 (1955), 218–244; III. *Math. Z.*, 63 (1955), 53–75; IV. *Amer. J. Math.*, 77 (1955), 429–452; V. *Bull. Soc. Math. France*, 84 (1956), 207–239; VI. *Amer. J. Math.*, 79 (1957), 331–388; VII. *Math. Ann.*, 134 (1957), 114–133; VIII. *Amer. J. Math.*, 80 (1958), 740–772.

[13]. Freudenthal, H., Die Topologie der Lieschen Gruppen als algebraisches Phänomen, I. *Ann. Math.*, 42 (1941), 1051–1074.

[14]. —— *Oktaven, Ausnahmegruppen und Oktavengeometrie.* Mathematisch Instituut der Rijksuniversiteit, Utrecht, 1951 (2e éd. 1960).

[15]. —— Beziehungen der E_7 und E_8 zur Oktavenebene. I–IX, *Proc. Kon. Ned. Akad. Wet.*, A 57 (1954), 218–230, 363–368; A 58 (1955), 151–157, 277–285; A 62 (1959), 165–179, 180–191, 192–201, 447–465, 466–474 (= *Indag. Math.*, 16, 17, 21).

[16]. Galois, E., *Œuvres mathématiques.* Gauthier-Villars, Paris, 1951.

[17]. Hertzig, D., Forms of algebraic groups. *Proc. Amer. Math. Soc.*, 12 (1961), 657–660.

[18]. JORDAN, C., *Traité des substitutions et des équations algébriques*. Gauthier-Villars, Paris, 1870.

[19]. KILLING, W., Die Zusammensetzung der stetigen endlichen Transformationsgruppen. I–IV. *Math. Ann.*, 31 (1888), 252–290; 33 (1889), 1–48; 34 (1889), 57–122; 36 (1890), 161–189.

[20]. LANG, S., Algebraic groups over finite fields. *Amer. J. Math.*, 78 (1956), 555–563.

[21]. —— *Introduction to Algebraic Geometry*. Interscience, New York, 1958.

[22]. —— *Abelian varieties*, Interscience, New York, 1959.

[23]. MATHIEU, E., Mémoire sur l'étude des fonctions de plusieurs quantités, sur la manière de les former et sur les substitutions qui les laissent invariables. *J. Math. Pures Appl.*, 6 (1861), 241–243.

[24]. —— Sur les fonctions cinq fois transitives de 24 quantités. *J. Math. Pures Appl.*, 18 (1873), 25–46.

[25]. MILLER, G. A., Sur plusieurs groupes simples. *Bull. Soc. Math. France*, 28 (1900), 266–267.

[26]. REE, R., On some simple groups defined by C. Chevalley. *Trans. Amer. Math. Soc.*, 84 (1957), 392–400.

[27]. —— A family of simple groups associated with the simple Lie algebra of type (F_4). *Amer. J. Math.*, 83 (1961), 401–420.

[28]. —— A family of simple groups associated with the simple Lie algebra of type (G_2). *Amer. J. Math.*, 83 (1961), 432–462.

[29]. ROSENFELD, B. A., Geometričeskaya interpretaciya kompaktnyh prostyh grupp Li klassa E (Interprétation géométrique des groupes de Lie de la classe E). *Dokl. Akad. Nauk S.S.S.R.*, 106 (1956), 600–603.

[30]. —— Einfache Lie-Gruppen und nichteuklidische Geometrien, *Coll. Alg. & top. found. geom.* (Utrecht, 1959) 135–155. Pergamon, Oxford, 1962.

[31]. SATAKE, I., On the theory of reductive algebraic groups over a perfect field. *J. Math. Soc. Japan.* (A paraître.)

[32]. SCHELLEKENS, G. L., On a hexagonic structure, *Proc. Kon. Ned. Akad. Wet.*, A 65 (1962). (= *Indag. Math.* 24),

[33]. SEGRE, B., On complete caps and ovaloids in three-dimensional Galois spaces of characteristic two. *Acta Arithmetica*, 5 (1959), 315–332.

[34]. SELIGMAN, G. B., On automorphisms of Lie algebras of classical type. I–III. *Trans. Amer. Math. Soc.*, 92 (1959), 430–448; 94 (1960), 452–482; 97 (1960), 286–316.

[35]. SERRE, J.-P., Cohomologie galoisienne des groupes algébriques linéaires, *Coll. Théorie des groupes (Groupes algébriques), Centre Belge Rech. Math.* Bruxelles, juin 1962. (A paraître.)

[36]. STEINBERG, R., Variations on a theme of Chevalley, *Pacific J. Math.*, 9 (1959), 875–891.

[37]. —— The simplicity of certain groups. *Pacific J. Math.*, 10 (1960), 1039–1041.

[38]. SUZUKI, M., A new type of simple groups of finite order. *Proc. Nat. Acad. Sci. U.S.A.*, 46 (1960), 868–870.

[39]. TITS, J., Le plan projectif des octaves et les groupes de Lie exceptionnels. *Acad. Roy. Belg., Bull. Cl. Sci.*, 39 (1953), 309–329.

[40]. —— Le plan projectif des octaves et les groupes exceptionnels E_6 et E_7. *Acad. Belg., Bull. Cl. Sci.*, 40 (1954), 29–40.

[41]. —— Sur certaines classes d'espaces homogènes de groupes de Lie. *Mém. Acad. Roy. Belg.*, 29 (3) (1955).

[42]. —— Sur la géométrie des R-espaces. *J. Math. Pures Appl.*, 36 (1957), 17–38.

[43]. —— Les groupes de Lie exceptionnels et leur interprétation géométrique. *Bull. Soc. Math. Belg.*, 8 (1956), 48–81.

[44]. —— Sur les analogues algébriques des groupes semi-simples complexes. *Coll. Alg. sup., Centre Belge Rech. Math.* Bruxelles, 1956, 261–289. Ceuterick, Louvain, 1957.

[45]. —— Les « formes réelles » des groupes de type E_6. *Séminaire Bourbaki*, exposé n° 162, février 1958. Secrétariat mathématique, Paris.

[46]. —— Sur la trialité et certains groupes qui s'en déduisent. *Publ. Math. (I.H.E.S.), Paris*, 2 (1959), 14–60.

[47]. —— Sur la classification des groupes algébriques semi-simples. *C.R. Acad. Sci. Paris*, 249 (1959), 1438–1440.

[48]. —— *Sur les groupes algébriques affins: théorèmes fondamentaux de structure; classification des groupes semi-simples et géométries associées.* Centro Int. Mat. Estivo (Saltino di Vallombrosa, 1959). Rome, 1960.

[49]. —— Groupes algébriques semi-simples et géométries associées. *Coll. Alg. & top. found. geom.* (Utrecht, 1959), 175–192. Pergamon, Oxford, 1962.

[50]. —— Les groupes simples de Suzuki et de Ree, *Séminaire Bourbaki*, exposé 210, décembre 1960. Secrétariat mathématique, Paris.

[51]. —— Ovoïdes et groupes de Suzuki, *Arch. Math.*, 13 (1962), 187–198.

[52]. —— Théorème de Bruhat et sous-groupes paraboliques. *C. R. Acad. Sci. Paris*, 254 (1962), 2910–2912.

[53]. —— Géométries polyédriques et groupes simples. *IIe Réunion de Math. d'Expression latine* (Florence, 1960), 64–86. Cremonese, Rome, 1962.

[54]. Weil, A., Algebras with involutions and the classical groups. *J. Indian Math. Soc.*, 24 (1960), 589–623.

[55]. Witt, E., Theorie der quadratischen Formen in beliebigen Körpern. *J. reine angew. Math.*, 176 (1936), 33–34.

[56]. —— Die 5-fach transitiven Gruppen von Mathieu. *Abh. Math. Sem. Hamburg*, 12 (1938), 256–264.

[57]. —— Spiegelungsgruppen und Aufzählung halbeinfacher Liescher Ringe. *Abh. Math. Sem. Hamburg*, 14 (1941), 289–322.

HALF-HOUR ADDRESSES

REMARKS ON FOUNDATIONS OF PROBABILITY

SEMANTICAL INTERPRETATION OF THE PROBABILITY OF FORMULAS

By JERZY ŁOŚ

Every Boolean algebra may be isomorphically represented as an algebra of sets, therefore every probability function on a Boolean algebra may be represented as a probability on some algebra of sets. This follows from the representation theory due to M. Stone.

If one considers the probabilities defined for formulas of a given formalism, which as is well known form a Boolean algebra, then such a representation seems to be insufficient, because it has no semantical meaning.

The following paper contains a representation of probability functions defined on formulas by means of semantical notions. The starting point of this representation is in the two different interpretations of probability of formulas, according to the syntactical structure of the formula. If the formula is the open one, i.e. no quantifiers occur in it, then by its probability we may understand the chance of finding out such elements in a given model which fulfil this formula. If the formula is a sentence, i.e. no free variables occur in it, then by its probability we may understand the chance of finding out a model (from among models of a given class) in which this sentence is true.

The main result obtained in this paper (Theorem 2) is the demonstration that, under very general conditions (which are always satisfied by elementary theories), the probability of sentences may be understood as a result of two consecutive drawings. First a model will be chosen at random from among models of a given class $\{\mathfrak{M}_t\}_{t \in T}$ following a given probability μ in set T; and then, having the so obtained model $\mathfrak{M}_t$, a sequence of elements is selected from it, following a probability ν_t. The probability $p(\alpha)$ of the formula α is equal to the probability of fulfilling α by a sequence of elements obtained in this way. More precisely this theorem states that we can represent the probability of formulas $p(\alpha)$ in the form

$$p(\alpha) = \int_T \nu_t(\mathfrak{M}_t(\alpha))\, d\mu(t),$$

where $\{\mathfrak{M}_t\}_{t \in T}$ is a class of models, μ a probability in T, ν_t a probability in the set of sequences of elements of the model $\mathfrak{M}_t$ and $\mathfrak{M}_t(\alpha)$ denotes the set of sequences which fulfils α in the model $\mathfrak{M}_t$.

It is not the aim of this paper to explain the importance of research on the probability of formulas. We would like to remark only that the theory of information yields more and more interest in such investigations.

1. Formulas, Boolean algebras and models

We shall here be interested in those formal theories in which only relational constants occur. Variables of higher orders (e.g. running over sets or classes of sets) may be allowed, but in such cases we shall assume that in the formulas each such variable is bounded by means of a quantifier. Only individual variables may be free in these formulas. The set of all formulas is denoted by S.

Let us denote the relational constants by r_i and the individual variables occurring in formulas in S by x_j. We shall assume that i runs over a set of natural numbers and j over the set of all natural numbers. This implies that the set S is countable.

By a sentence is meant a formula in which no free variable occurs. The set of sentences will be denoted by Z; it is a subset of S.

Let Cn be the operation of the usual (syntactical) consequence in the set S. A subset X of S with $X = Cn(X)$ is a system, a subset X with $Cn(X) \neq S$ is a consistent set, a maximal consistent system is a complete system.

Let us remember the following properties of the set Z:

(i) The set Z is closed with respect to the propositional operations such as implication ($\rightarrow$), disjunction ($\vee$), conjunction ($\wedge$), equivalence ($\leftrightarrow$) and negation ($\sim$).

(ii) For every system $X, Cn(X \cap Z) = X$.

As is well known, the set of formulas form a Boolean algebra with disjunction, conjunction and negation as Boolean operations. Instead of equality we have to accept in S the equivalence with respect to the set $Cn(O)$, i.e. two formulas α and β are considered to be equal if $\alpha \leftrightarrow \beta$ belongs to $Cn(O)$. In this algebra $\alpha \cap \beta = O$ means that $\sim(\alpha \wedge \beta)$ belongs to $Cn(O)$. It follows from (i) that Z is a subalgebra of S and so it is called the algebra of sentences.

By a model for S we shall mean every sequence $\mathfrak{M} = \langle A, R_1, R_2, \ldots \rangle$ where A is non-empty set, and R_i relations in A such that it is possible to interpret by them constants r_i. It is clear what is meant by the statement that the sequence $\langle a_1, a_2, \ldots \rangle$ of the elements of A fulfils the formula α. Let us remark only that the individual variables have natural numbers as indices and therefore we use here an infinite sequence of elements of A. Moreover, the variables of higher orders (if they occur in formulas in S) have to be interpreted in an absolute sense, e.g. a set variable runs over all subset of A.

Let us denote by $\mathcal{A}$ the set of all sequences $\langle a_1, a_2, \ldots \rangle$ of elements of A and, for a given α in S, by $\mathfrak{M}(\alpha)$ the set of all sequences in $\mathcal{A}$ which fulfils α. $\mathfrak{M}(\alpha)$ is a homomorphism of S into an algebra of subsets of $\mathcal{A}$.

By $S(\mathfrak{M})$ we denote the set of all formulas "true" in $\mathfrak{M}$, i.e., the set of α's with $\mathfrak{M}(\alpha) = \mathcal{A}$. If for a system X there exists a model $\mathfrak{M}$ such that X is contained in $S(\mathfrak{M})$, then X is called an ω-consistent, in other cases an ω-inconsistent system. If for some class of systems $\{\mathfrak{M}_t\}_{t \in T}$, $X = \bigcap_{t \in T} S(\mathfrak{M}_t)$, then X is called ω-regular. If for every complete system Y including the given system X there exists a model $\mathfrak{M}$ with $S(\mathfrak{M}) = Y$, then X is called strongly ω-regular. If S is an elementary class, i.e. no variables of higher orders occur in the formulas of S, then every system in S is strongly ω-regular.

2. Probability functions

By a probability function or simply probability, is meant a real function p defined on a Boolean algebra B, such that $0 \leqslant p(\alpha) \leqslant 1$ and $p(a \cup b) = p(a) + p(b)$for every pair a, b of elements in B with $a \cap b = 0$.

If B is a σ-algebra and for evevery sequence of pairwise disjoint elements a_i we have $p(\bigcup_{i=1}^{\infty} a_i) = \sum_{i=1}^{\infty} p(a_i)$ then p is called σ-additive.

Let B^* be a σ-algebra and B a subalgebra of B^* such that B^* is the least σ-algebra containing B. If p is a probability on B and $p(a_1 \cap a_2 \cap \ldots \cap a_n) \to 0$ whenever $\bigcap_{i=1}^{\infty} a_i = 0$, then p may be extended to a σ-additive probability function on B^*. This result is very well known, as Kolmogoroff's theorem.

For the Boolean algebra of formulas and its subalgebras the definition of the probability function is obviously applicable. The only restriction which follows from the definition of such algebras is that for two formulas α and β equivalent with respect to the set $Cn(O)$ we must have $p(\alpha) = p(\beta)$.

A probability function p on the algebra of formulas S is called continuous, if the following two conditions hold:

There exists an ω-regular system X_0, such that $p(\alpha) = 1$ for every α in X_0.

If for some $\varepsilon > 0$ and some system X, $p(\zeta) \geqslant \varepsilon$ for every sentence ζ belonging to X, then X is an ω-consistent system.

3. Semantical introduction of probability

Let $\mathfrak{M}$ be a model and A its set of elements. By $\mathcal{A}$ we denote, as before, the set of sequences of elements from A. A probability function ν defined for an algebra of subsets of $\mathcal{A}$ is called a probability in the model $\mathfrak{M}$ if ν is defined for all sets $\mathfrak{M}(\alpha)$ with α in S. Suppose that $\{\mathfrak{M}_t\}_{t \in T}$ is a family of models and that for every t in T, ν_t is a probability in $\mathfrak{M}_t$. Suppose, moreover, that there is a σ-additive probability function μ on a σ-algebra of subsets of T, such that all functions $f_a(t) = v_t(\mathfrak{M}_t(\alpha))$, with α in S, are μ-measurable.

THEOREM 1. *Under the above assumptions, the function*

$$p(\alpha) = \int_T f_\alpha(t)\, d\mu(t)$$

is a probability on S.

Proof. $0 \leqslant p(\alpha) \leqslant 1$ follows from the assumption that ν_t and μ are probability functions. If $\alpha \cap \beta = 0$, then $\sim(\alpha \wedge \beta)$ belongs to $Cn(0)$. Therefore $\mathfrak{M}_t(\alpha) \cap \mathfrak{M}_t(\beta) = 0$ for every t. Since $\mathfrak{M}_t$ is a homomorphism hence:

$$\begin{aligned} p(\alpha \cup \beta) &= \int_T f_{\alpha \vee \beta}(t)\, d\mu(t) = \int_T \nu_t(\mathfrak{M}_t(\alpha \vee \beta))\, d\mu(t) \\ &= \int_T \nu_t(\mathfrak{M}_t(\alpha) \cup \mathfrak{M}_t(\beta))\, d\mu(t) = \int_T \nu_t(\mathfrak{M}_t(\alpha)) + \nu_t(\mathfrak{M}_t(\beta))\, d\mu(t) \\ &= \int_T \nu_t(\mathfrak{M}_t(\alpha))\, d\mu(t) + \int_T \nu_t(\mathfrak{M}_t(\beta))\, d\mu(t) \\ &= \int_T f_\alpha(t)\, d\mu(t) + \int_T f_\beta(t)\, d\mu(t) = p(\alpha) + p(\beta). \end{aligned}$$

This theorem shows the possibility of introducing probability functions in the algebra S by means of the probability functions on models and a probability function in the set of models. In the following section we shall prove that every continuous probability function on S may be represented in the form of an integral as in theorem 1.

4. Representation theorem

Let p be a continuous probability on S, X_0 an ω-regular system with $p(\zeta)=1$ for all sentences ζ in X_0 and $\{\mathfrak{M}_t\}_{t\in T}$ a class of models such that $X_0=\bigcap_{t\in T} S(\mathfrak{M}_t)$ and moreover such that for every system X containing X_0 with $p(\zeta)\geqslant\varepsilon$ for every sentence ζ in X and a suitable $\varepsilon>0$ depending on X, there exists a t in T such that $X\subset S(\mathfrak{M}_t)$. The existence of such a X_0 and $\{\mathfrak{M}_t\}_{t\in T}$ follows from the assumption on continuity of p.

Let for ζ in $Z, T(\zeta)$ denotes the set of t in T with ζ belonging to $S(\mathfrak{M}_t)$. $T(\zeta)$ maps the algebra of sentences Z homomorphically into the algebra of subsets of T.

Let us set $\mu(T(\zeta))=p(\zeta)$. This defines a probability function μ on the algebra of subsets of T of the form $T(\zeta)$. We shall prove that μ may be extended to a σ-additive probability function. Let $\mu(T(\zeta_1\wedge\ \dots\ \wedge\zeta n))\geqslant\varepsilon>0$ for every n. Let us form the system $Cn(X_0\cup(\zeta_1,\zeta_2,\dots))=X_1$. Obviously for every ζ in $X_1, \mu(T(\zeta))=p(\zeta)\geqslant\varepsilon$ and, therefore, there exists a t_1 in T such that $X_1\subset S(M_{t_1})$. It follows that t_1 belongs to $T(\zeta_n)$, for every n Finally $\bigcap_{n=1}^{\infty}T(\zeta_n)\neq 0$.

Let us, applying Kolmogoroff's theorem, extend μ to a σ-additive probability, denoting this extension by the same letter μ.

Now, for every α in S, let μ_α be a σ-additive probability function such that

$$\mu_\alpha(T(\zeta))=p(\zeta\wedge\alpha)$$

for every ζ in Z.

Such a probability does exist, because the equation above defines a probability on the sets of the form $T(\zeta)$, which may be extended (by the same argument as for μ) to a σ-additive probability.

Every μ_α is obviously absolutely continuous with respect to μ because $\mu_\alpha(T(\zeta))\leqslant\mu(T(\zeta))$ for every α in S and ζ in Z. From the well-known Radon-Nikodym theorem it follows that for every α in S there exists a function $f_\alpha(t)$ such that

$$\mu_\alpha(T(\zeta))=\int_{T(\zeta)} f_\alpha(t)\,d\mu(t).$$

Suppose α and β are disjoint, since $\mu_\alpha(T(\zeta))+\mu_\beta(T(\zeta))=p(\zeta\wedge\alpha)+p(\zeta\wedge\beta)=p(\zeta\wedge(\alpha\vee\beta))=\mu_{\alpha\vee\beta}(T(\zeta))$, hence

$$\int_{T\ \zeta)} f_\alpha(t)\,d\mu(t)+\int_{T(\zeta)} f_\beta(t)\,d\mu(t)=\int_{T(\zeta)} f_\alpha(t)+f_\beta(t)\,d\mu(t)$$

$$=\int_{T(\zeta)} f_{\alpha\vee\beta}(t)\,d\mu(t).$$

It shows that $f_\alpha + f_\beta = f_{\alpha \vee \beta}$ almost everywhere in the sense of μ. As we have only countably many α's and β's in S we can improve the functions f_α to have the equations $f_\alpha + f_\beta = f_{\alpha \vee \beta}$ holding everywhere in T for disjoint α's and β's.

Now, if we set for every t in T and α in S $\nu_t(\mathfrak{M}_t(\alpha)) = f_\alpha(t)$ we obtain a probability function on the subset $\mathfrak{M}_t(\alpha)$ of $\mathcal{A}_t$, such that

$$p(\zeta \wedge \alpha) = \int_{T(\zeta)} \nu_t(\mathfrak{M}_t(\alpha))\, d\mu(t)$$

For ζ in X_0 we have $p(\zeta \wedge \alpha) = p(\alpha)$, $T(\zeta) = T$ and therefore

$$p(\alpha) = \int_T \nu_t(\mathfrak{M}_t(\alpha))\, d\mu(t).$$

We have proved the following representation theorem:

THEOREM 2. *Every continuous probability on the algebra of sentences S may be represented in form of an integral*

$$p(\alpha) = \int_T \nu_t(\mathfrak{M}_t(\alpha))\, d\mu(t),$$

where $\mathfrak{M}_t$ are suitable models, ν_t probabilities in these models, and μ a σ-additive probability function in the set of models.

MODULES OVER UNRAMIFIED REGULAR LOCAL RINGS

By MAURICE AUSLANDER(1)

A ring R is said to be a local ring if it is a commutative, noetherian ring with unit with only one maximal ideal which we shall denote by $\mathfrak{m}$. Krull introduced in [8] the idea of studying a local ring R by means of its associated graded ring $G(R)=\sum_{i=0}^{\infty}\mathfrak{m}^i/\mathfrak{m}^{i+1}$. In particular, he was especially interested in the case when $G(R)\approx R/\mathfrak{m}[X_1,\ldots,X_n]$ (where the X_i are algebraically independent over $R/\mathfrak{m}$), the case we now denote by saying that R is a regular local ring. Zariski in [12] pointed out the important role that regular local rings play in algebraic geometry over a field and established the following facts about the regular local rings which occur in this context:

(*a*) If R is a regular local ring and $\mathfrak{p}$ is a prime ideal, then $R_{\mathfrak{p}}$ is also a regular local ring.

(*b*) If R is a regular local ring, then R is a unique factorization domain.

After the appearence of Cartan-Eilenberg's systematic approach to homological algebra [7], it was conjectured by many and established by Serre [11] that regular local rings can be completely characterized in homological terms as those local rings such that all their finitely generated modules have finite free resolutions. As a result of the work of Auslander-Buchsbaum, Nagata and Serre (see [4, 5], [9, 10], [11]) in the application of homological methods to the study of arbitrary regular local rings, the results cited above for geometric regular local rings are now known to hold for arbitrary regular local rings.

While the above cited work has concerned itself mainly with what might be considered the ideal theoretic aspects of regular local rings, more recent results have shown that it is probably desirable to have a more complete module theory for regular local rings than now exists. For example, in connection with the study of the Brauer group of regular local rings, Auslander-Goldman were lead to establish in [6] the following module theoretic generalization of the fact that a regular local ring is a unique factorization domain. Let M be a finitely generated, reflexive module over the regular local ring R (i.e. $M\approx \mathrm{Hom}_R(\mathrm{Hom}_R(M,R),\ R)$). If $\mathrm{Hom}_R(M,M)$ considered as an R-module is a free R-module, then M is a free R-module.

More recently, Auslander in [3] established and showed that the following result can be viewed as a generalization of the purity of the branch locus theorem in algebraic geometry. Let M be a finitely generated reflexive module over the regular local ring R. If $\mathrm{Hom}_R(M,M)$ is isomorphic to a direct sum of copies of M, then M is a free R-module.

Proceeding on the premise that the module theory of regular local rings will gradually assume more importance, both in the abstract theory of regular local rings as well as in applications to other fields, we will devote

(1) Some of this work was done while the author held an N.S.F. Senior Postdoctoral Fellowship and a Sloan Foundation Fellowship.

the rest of this report to a discussion of certain recent results obtained by the author concerning finitely generated modules over unramified regular local rings. Since complete proofs of these results have appeared, or will shortly appear in the literature [1, 2], we will only concern ourselves here with the presentation of the results themselves.

We assume from now on that R is an unramified regular local ring, i.e. if the characteristic of the residue class field is $p \neq 0$ and p is in $\mathfrak{m}$, then p is not in $\mathfrak{m}^2$. It can easily be seen that if R is an unramified regular local ring, then $R_\mathfrak{p}$ is also an unramified regular local ring for all prime ideals $\mathfrak{p}$ in R. We will also assume that all R-modules are finitely generated.

The reason for assuming that R is an unramified regular local and not an arbitrary regular local ring, is that we only know how to establish the following result for that smaller class of regular local rings.

THEOREM 1. *Let A and B be R-modules such that* $\mathrm{Tor}_i^R(A,B)=0$ *for some* $i \geqslant 0$. *Then* $\mathrm{Tor}_j^R(A,B)=0$ *for all* $j \geqslant i$.

If this result can be established for arbitrary regular local rings, then the rest of the results, which are more or less direct consequences of this one fact, will also hold for arbitrary regular local rings since this is the only place that the assumption that R is an unramified regular local ring and not an arbitrary regular local ring is needed.

If one defines the support of a module M to be the set of all prime ideals $\mathfrak{p}$ in R such that $M_\mathfrak{p} \neq 0$ (notation: Supp M), then Theorem 1 yields a nice formula for computing the Supp $\mathrm{Tor}_i^R(A, B)$. Namely,

THEOREM 2. *A prime ideal p is in the* Supp $\mathrm{Tor}_i(A,B)$ *if and only if there is a prime ideal $\mathfrak{q}$ contained in $\mathfrak{p}$ such that*

$$hd_{R_\mathfrak{q}} A_\mathfrak{q} + hd_{R_\mathfrak{q}} B_\mathfrak{q} \geqslant \dim R_\mathfrak{q} + i,$$

where $\dim R_\mathfrak{q}$ is the Krull dimension of the local ring $R_\mathfrak{q}$ and $hd_{R_\mathfrak{q}} A_\mathfrak{q}$ means the length of the shortest projective resolution of $A_\mathfrak{q}$ as an $R_\mathfrak{q}$-module (it being understood that the $hd_R O = -\infty$ and $(-\infty)+(-\infty)=-\infty$ and $n+(-\infty)=-\infty$ for all integers n).

This suggests introducing the following definition. Two modules A and B are said to have the same *dimension type* if and only if the $hd_{R_\mathfrak{p}} A_\mathfrak{p} = hd_{R_\mathfrak{p}} B_\mathfrak{p}$ for all prime ideals $\mathfrak{p}$ in R. We will denote the equivalence class of all modules with the same dimension type as A by $[A]$. It is also convenient to introduce an order relation in the equivalence classes $[A]$ by defining $[A] \geqslant [B]$ if and only if the $hd_{R_\mathfrak{p}} A_\mathfrak{p} \geqslant hd_{R_\mathfrak{p}} B_\mathfrak{p}$ for all prime ideals $\mathfrak{p}$ in R. It is then an immediate consequence of Theorem 2 that the Supp $\mathrm{Tor}_i^R(A,B)$ depends only on the dimension types of A and B. Also we have

THEOREM 3. $[A] \geqslant [B]$ *if and only if the* Supp $\mathrm{Tor}_i^R(A,C) \supset$ Supp $\mathrm{Tor}_i^R(B,C)$ *for all* $i \geqslant 0$ *and all R-modules C.*

Proceeding in a somewhat different direction, we obtain the following novel criteria for a module to be a free module.

THEOREM 4. *Suppose that the* $\dim R = n > 0$ *and that A is an R-module. Then*

(*a*) *A is a free R-module if and only if the tensor product of A with itself n times is torsion free.*

(*b*) *If* $\mathrm{Hom}_R(A,R)\neq 0$, *then* A *is a free module if and only if* $A\otimes_R A\otimes_R \mathrm{Hom}(A,R)$ *is torsion free.*

(*c*) *If* A *satisfies the following conditions:*

(i) $hd\ A = hd\ \mathrm{Hom}(A,R)$

(ii) $A\otimes \mathrm{Hom}(A,R)$ *is torsion free*

(iii) $A_{\mathfrak{p}}$ *is* $R_{\mathfrak{p}}$*-free for all non-maximal prime ideals* $\mathfrak{p}$ *in* R, *then the hd* $A=(n-1)/2$ *or* 0. *In particular if* n *is even then* A *is free.*

It should be observed that each of the results given in Theorem 4 are the best possible. For if $n\geqslant 3$, then there exist modules A such that the tensor product of A with itself $(n-1)$ times is torsion free but A is not free. As far as (*c*) is concerned, in case $n>1$ and odd, there always exist modules A satisfying the hypothesis of (*c*) with the $hd\ A=(n-1)/2$; which shows that the result (*b*) is the best possible as well as indicating that there seem to be some differences in the module theory of unramified regular local ring depending on whether the dimension of the ring is odd or even.

We now devote the rest of this report to a discussion of some of the information concerning the Ext functor which can be derived from the above discussion of the Tor functor.

If A and B are R-modules, there exists a natural homomorphism φ: $\mathrm{Hom}(A,R)\otimes_R B\to \mathrm{Hom}(A,B)$ given by $\varphi(f\otimes b)(a)=f(a)b$. This can be generalized to give natural homomorphisms $\mathrm{Ext}^i(A,R)\otimes_R B\to \mathrm{Ext}^i(A,B)$ which are closely related to the usual Yoneda product. A sort of universal coefficient theorem tells us that for a fixed A and each $i\geqslant 0$ there exist modules T_{i+1} such that we have exact sequences

$$\mathrm{Tor}_2^R(\mathrm{T}_{i+1}, B)\to \mathrm{Ext}_R^i(A, R)\otimes_R B\to \mathrm{Ext}_R^i(A, B)\to \mathrm{Tor}_1^R(T_{i+1}, B)\to 0.$$

Combining this observation with the foregoing results we obtain:

THEOREM 5. *If* $j\geqslant 0$ *and* $\mathrm{Ext}_R^j(A, R)\otimes_R B\to \mathrm{Ext}_R^j(A, B)\to 0$ *is exact, then*

(*a*) $\mathrm{Ext}_R^j(A, R)\otimes_R B=\mathrm{Ext}_R^j(A, B)$.

(*b*) *If* $[B]\geqslant [C]$, *then* $\mathrm{Ext}_R^j(A, R)\otimes_R C=\mathrm{Ext}_R^j(A, C)$.

(*c*) *There exists an exact sequence*

$$\mathrm{Ext}_R^{j-1}(A, R)\otimes B\to \mathrm{Ext}_R^{j-1}(A, B)\to \mathrm{Tor}_1^R(\mathrm{Ext}_R^j(A, R), B)\to 0.$$

(*d*) *If* $[A]=[B]$, *then the* $hd_R A=hd_R B\leqslant j$ *and one has equality if and only if* $\mathrm{Ext}_R^j(A, B)\neq 0$.

COROLLARY 6. *If* $B\neq 0$ *and* $\mathrm{Ext}_R^j(A, B)=0$ *for some* $j\geqslant 0$, *then*

(*a*) $\mathrm{Ext}_R^j(A, R)=0$.

(*b*) *If* $[B]\geqslant [C]$, *then* $\mathrm{Ext}_R^j(A, C)=0$ *and* $\mathrm{Ext}_R^{j-1}(A, R)\otimes_R C=\mathrm{Ext}_R^{j-1}(A, C)$.

(*c*) *If* $[A]=[B]$, *then the* $hd_R A=hd_R B<j$.

Finally these results can be combined to show that for a fixed A and i, the Supp $\mathrm{Ext}_R^i(A, B)$ depends only on the dimension type of B. However, there exist examples which show that for a fixed B and i the Supp $\mathrm{Ext}_R^i(A, B)$ does not depend solely on the dimension type of A.

REFERENCES

[1]. AUSLANDER, M., Modules over unramified regular local rings. *Illinois J. Math.*, 5 (1961), 631–647.

[2]. —— Modules over unramified regular local rings. II. *Illinois J. Math.* (To appear.)

[3]. —— On the purity of the branch locus. *Amer. J. Math.*, 84 (1962), 116–125.

[4]. AUSLANDER, M. & BUCHSBAUM, D., Homological dimension in local rings. *Trans. Amer. Math. Soc.*, 85 (1957), 390–405.

[5]. —— Unique factorization in regular local rings. *Proc. Nat. Acad. Sci. U.S.A.*, 45 (1959), 733–734.

[6]. AUSLANDER, M. & GOLDMAN, O., Maximal orders. *Trans. Amer. Math. Soc.*, 97 (1960), 1–24.

[7]. CARTAN, H. & EILENBERG, S., *Homological Algebra*. Princeton University Press, 1956.

[8]. KRULL, W., Dimensiontheorie in Stellenringen. *J. reine angew. Math.*, 179 (1938), 204–226.

[9]. NAGATA, M., A general theory of algebraic geometry over dedekind rings. II. *Amer. J. Math.*, 80 (1958), 382–420.

[10]. —— The theory of multiplicity in general local rings. *Proc. Int. Symp. Alg. Number Theory*. Tokyo, 1956, 191–226.

[11]. SERRE, J.-P., Sur la dimension des anneaux et des modules noethériens. *Proc. Int. Symp. Alg. Number Theory*. Tokyo, 1956, 175–189.

[12]. ZARISKI, O., The concept of a simple point of an abstract algebraic variety. *Trans. Amer. Math. Soc.*, 62 (1947), 1–52.

ARITHMETIC ON AN ELLIPTIC CURVE

By J. W. S. CASSELS

1. One of the fundamental problems of number theory is this: given an algebraic variety V defined over an algebraic numberfield(1) k to find all the points on V defined over k and to investigate their structure and properties. For example V might be a variety in affine n-dimensional space given by a set of equations

$$f_j(x_1, \ldots, x_n) = 0 \quad (1 \leqslant j \leqslant r), \tag{1.1}$$

where the f_j are polynomials with coefficients in the groundfield k, and points defined over k are just the sets of values of $(x_1, \ldots, x_n)$ in k which satisfy the equations.

In this talk we shall be concerned almost entirely with curves and it is convenient technically to use the word point to mean a place of the corresponding functionfield. Alternatively we could confine attention to complete curves without multiple points, in which case the points on the curve in the naive sense are precisely the points in this sense too.

Let V be an algebraic variety defined over the algebraic numberfield k and let K be any field containing k. A trivial necessary condition for the existence of points on V defined over k is that there should be points defined over K. In particular, if we let K run through all the completions k_v of k with respect to all its valuations v we get a set of necessary conditions for the existence of a point defined over k. Roughly speaking this is the same as saying that the equations defining V must (i) have solutions when regarded as congruences to any modulus and (ii) have real solutions for any embedding of k in the real field. We shall speak of the behaviour of V in the local fields k_v as its local behaviour and, in contrast, we shall speak of its behaviour in k as global. In this terminology we have just said that a necessary condition for the existence of points on V globally is the existence of points everywhere locally.

For curves of genus 0 there is now a completely satisfactory state of affairs. The existence of points everywhere locally is not merely necessary but also sufficient for the existence of points globally. When this criterion is satisfied the curve is birationally equivalent over k to the projective straight line, i.e. can be parametrized by a single variable t in such a way that the points defined over k are just those for which the parameter t is in k (or is ∞). Further, in any given case one can decide whether the criterion is satisfied and, if so, set up the parametrization, all in a finite number of steps. Hence the theory of points on curves of genus 0 defined over an algebraic number field is devoid of interest because there is nothing left to do. (See [36] for details.)

For curves of genus greater than 1 the situation is quite the opposite. There is only the long-standing conjecture of Mordell that there is only

(1) By an algebraic numberfield I shall mean an algebraic extension of the rationals of finite degree.

a finite number of points defined over an algebraic numberfield on such a curve : but little progress seems to have been made with it.(1)

2. There remain the curves of genus 1 to which the rest of this talk will be devoted. Here the situation is interesting in that some general results are known, there are plausible conjectures which still resist proof, and it is likely that much remains to be discovered.

A curve C of genus 1 defined over a field k which has on it a point $\mathfrak{a}$ defined over k is an abelian variety of dimension 1 defined over k. The points on C, whether defined over k or not, have the structure of an abelian group(2) for which $\mathfrak{a}$ is the zero element. This group structure is defined(3) over k. For example, C could be the curve in 3-dimensional projective space given by the equations

$$\left.\begin{aligned} x_1^2 &= x_2^2 - x_4^2 \\ x_3^2 &= x_2^2 + x_4^2 \end{aligned}\right\} \tag{2.1}$$

in homogeneous coordinates (x_1, x_2, x_3, x_4). If $\mathfrak{o} = (1, 1, 1, 0)$, the coordinates of the sum of $\mathfrak{x} = (x_1, x_2, x_3, x_4)$ and $\mathfrak{y} = (y_1, y_2, y_3, y_4)$ is $\mathfrak{z} = (z_1, z_2, z_3, z_4)$, where

$$\left.\begin{aligned} z_1 &= y_1 y_4 x_2 x_3 - x_1 x_4 y_2 y_3, \\ z_2 &= y_2 y_4 x_1 x_3 - x_2 x_4 y_1 y_3, \\ z_3 &= y_3 y_4 x_1 x_2 - x_3 x_4 y_1 y_2, \\ z_4 &= y_4^2 x_1^2 - x_4^2 y_1^2 = y_4^2 x_2^2 - x_4^2 y_2^2 = y_4^2 x_3^2 - x_4^2 y_3^2. \end{aligned}\right\} \tag{2.2}$$

Under these circumstances the set of all points of C defined over k form a group, which we shall denote by $\mathfrak{G}$. When the groundfield is the rational field it was shown by Mordell [19] that $\mathfrak{G}$ is finitely generated.(4) This was extended by Weil [39] to all algebraic numberfields as groundfield and to abelian varieties of all dimension. The proofs are not, however, constructive : they give an upper bound for the number of generators of $\mathfrak{G}$ but give no procedure at all for finding them. I shall come back to this point later.

---

(1) Manin [17] has recently obtained interesting general results for the corresponding problem when the groundfield is a function-field, but his methods seem special to this case. Of course there is the Mordell-Weil theorem about rational sets of g points on a curve of genus g but when $g \neq 1$ this tells us nothing in general about the rational points. Chabauty [44] has however shown that there are only finitely many points on the curve if the Mordell-Weil group has fewer than g generators.

(2) If C is written in Weierstrass' normal form

$$y^2 = 4x^3 - g_2 x - g_3$$

with, say, the complex field as groundfield, then the addition is just the addition of the Weierstrass parameters (modulo the periods) when o is taken as the point at infinity.

(3) That is, the coordinates of the sum $\mathfrak{x} + \mathfrak{y}$ of two points $\mathfrak{x}, \mathfrak{y}$ and of $-\mathfrak{x}$ are rational functions of the coordinates of $\mathfrak{x}, \mathfrak{y}$ with coefficients in k.

(4) Thus it was conjectured by Leonardo Pisano about $+1220$ and proved by Fermat (in his marginal notes in his copy of Bachet's edition of *Diophantus' Arithmetica*) that the only rational points on (2.1) are the four points with $x_4 = 0$. The group $\mathfrak{G}$ for (2.1) with k as the rational field is thus the direct product of two cyclic groups of order 2.

The points of finite order in $\mathfrak{G}$ can, however, always be determined. When the groundfield is the rationals and when C is taken in the form

$$y^2 = x^3 - Ax - B \quad (A, B \text{ rational integers}) \tag{23}$$

(which is no real restriction) it was shown by Nagell [19] that rational points of finite order must have integral coordinates x, y and that y^2 is either 0 or a divisor of the discriminant $4A^3 - 27B^2$. There are thus only a finite number of candidates to look at. When k is an algebraic numberfield the points of finite order must satisfy similar arithmetic conditions (see, for example, Cassels [4]) but these are not now sufficient to show that the number of points of finite order is finite. However, it was shown by Lutz [16] that the group $\mathfrak{G}_v$ of points defined over the completion k_v of k with respect to any nonarchimedean valuation v contains a subgroup of finite index isomorphic to the additive group of elements of(1) k_v. In particular, even $\mathfrak{G}_v$ contains only a finite number of elements of finite order. The process is constructive(2) and allows the points of finite order in both $\mathfrak{G}$ and $\mathfrak{G}_v$ to be determined.

3. A curve $\mathcal{D}$ of genus 1 defined over an algebraic numberfield k does not, of course, necessarily have on it a point defined over k. There is, however, associated with $\mathcal{D}$ another curve C of genus 1 defined over k which does have on it a point defined over k, namely its jacobian curve. The jacobian C is determined(3) by $\mathcal{D}$ up to algebraic equivalence defined over k, so is unique for our purpose. Further, $\mathcal{D}$ has the structure of a principal homogeneous space(4) over the group C. That means that if $\mathfrak{X}$, $\mathfrak{x}$ are points of $\mathcal{D}$, C respectively then there is a uniquely determined point $\mu(\mathfrak{X}, \mathfrak{x})$ on $\mathcal{D}$ so that

(i) any two of the three points $\mathfrak{X}$, $\mathfrak{x}$, $\mu(\mathfrak{X}, \mathfrak{x})$ determines the third uniquely by equations with coefficients in k.

(ii) $$\mu\{\mu(\mathfrak{X}, \mathfrak{x}_1), \mathfrak{x}_2\} = \mu(\mathfrak{X}, \mathfrak{x}_1 + \mathfrak{x}_2), \tag{3.1}$$

where $\mathfrak{x}_1, \mathfrak{x}_2$ are any two points on C and $\mathfrak{x}_1 + \mathfrak{x}_2$ is their sum on C.

For example if l, m, n are any nonzero rational numbers the curve $\mathcal{D}$ given in 3-dimensional projective space by the equations

$$\left.\begin{aligned} mnX_1^2 &= nlX_2^2 - X_4^2, \\ lmX_4^2 &= nlX_2^2 + X_4^2, \end{aligned}\right\} \tag{3.2}$$

will not, in general, have a rational point. The jacobian C of $\mathcal{D}$ is just the curve given by (2.1). If $\mathfrak{X} = (X_1, X_2, X_3, X_4)$, $\mathfrak{x} = (x_1, x_2, x_3, x_4)$ are points of $\mathcal{D}$, C, respectively, then we can put

(1) In fact (cf. Weil [40]) the points defined over k_v in a sufficiently small v-adic neighbourhood of o can be parametrized by an analogue of the Weierstrass $\wp$-function. For this analogue the powerseries are formally the same but convergence is, of course, in the v-adic sense. It would not be reasonable to expect the parametrization to extend to all points defined over k_v, if only because no function in k_v defined by a powerseries can be periodic. Mattuck [18] has generalized all this to abelian varieties of any dimension.

(2) Schinzel reminded me during the Congress that Mahler [45] has given another effective way of determining the points of finite order on $\mathfrak{G}$ which works when k is an algebraic numberfield. It uses the machinery of the infinite descent.

(3) I think that the first proof of this was given by Châtelet [11]. For a more general theory see Lang [13].

(4) Principal homogeneous spaces over other algebraic groups have interesting arithmetical properties, see for example the expository paper of Lang 14] and Borel [3], and the lectures of Borel and Kneser at this conference.

$$\left.\begin{aligned}\mu(\mathfrak{X},\mathfrak{x}) = (Y_1, Y_2, Y_3, Y_4), \text{ where}\\ Y_1 &= X_1X_4x_2x_3 - lx_1x_4X_2X_3,\\ Y_2 &= X_2X_4x_1x_3 - mx_2x_4X_1X_3,\\ Y_3 &= X_3X_4x_1x_2 - nx_3x_4X_1X_2,\\ Y_4 &= X_4^2x_1^2 - mnx_4^2X_1^2 = X_4^2x_2^2 - nlx_4^2X_2^2 = X_4^2x_3^2 - lmx_4^2X_3^2.\end{aligned}\right\} \quad (3.3)$$

Of course in general the algebraic details are not as simple as in the example.

Let us now revert to the general case. The structure of homogeneous space on $\mathcal{D}$ is not unique. If μ is one, then another is μ', where

$$\mu'(\mathfrak{X},\mathfrak{x}) = \mu(\mathfrak{X},-\mathfrak{x}).$$

Usually this is the only way in which a second structure can occur, but there can be 4 or 6 structures in the special cases when C has "complex multiplication" by 4th or 6th roots of 1.

We shall work not with the principal homogeneous spaces themselves but with classes of principal homogeneous spaces. Two principal homogeneous spaces $(\mathcal{D}_1,\mu_1)$ and $(\mathcal{D}_2,\mu_2)$ over the same jacobian C are in the same class if there is a birational equivalence defined over k between $\mathcal{D}_1,\mathcal{D}_2$ which takes μ_1 into μ_2 in the obvious way.

The advantage of introducing classes of principal homogeneous spaces is that they naturally form a group, whereas there is no obvious structure on the equivalence classes of curves $\mathcal{D}$ with the same jacobian C(1). This group $WC = WC(\mathsf{C},k)$ is a torsion group. Its group structure is easiest explained in terms of a little homological algebra. Let $(\mathcal{D},\mu)$ be a homogeneous space for the jacobian C, all defined over k. There certainly will be some point $\mathfrak{A}$ on the curve $\mathcal{D}$ defined over the algebraic closure $\bar{k}$ of k. Then $\sigma\mathfrak{A}$ is also a point on $\mathcal{D}$, where σ is any element of the galois group Γ of $\bar{k}/k$. By the structure of principal homogeneous space there is a unique point $\mathfrak{a}_\sigma$ of C defined over $\bar{k}$ such that

$$\sigma\mathfrak{A} = \mu(\mathfrak{A}, \mathfrak{a}_\sigma). \quad (3.4)$$

It is readily verified that $\mathfrak{a}_\sigma$ is a cocycle with values in the group $\bar{\mathfrak{G}}$ of points on C defined over $\bar{k}$. A more detailed examination shows that this sets up an isomorphism between the group WC of classes of principal homogeneous spaces and a cohomology group $H^1(\Gamma,\bar{\mathfrak{G}})$. It is not quite the ordinary cohomology group because the point $\mathfrak{A}$ is defined already over a finite extension of k, which restricts the type of cocycle that can occur,(2) but it has the right sort of formal properties.(3) The natural group structure of $H^1(\Gamma,\bar{\mathfrak{G}})$ gives the group structure of WC.

(1) The situation is rather like that in elementary number theory. It is easier to study the classes of ideals in a ring $\mathbf{Z}[d^{\frac{1}{2}}]$ than to study the classes of binary quadratic forms of discriminant d because the former have a natural group structure and the latter do not. In general one class of quadratic forms corresponds to two classes of ideals.

(2) $\mathfrak{A}_\sigma$ depends only on the action of σ on a finite extension of k (depending on the cocycle). One way of expressing this is to say that one considers only continuous cocycles when Γ has the Krull and $\mathfrak{G}$ the discrete topology.

(3) For all this see Lang and Tate [15] or Šafarevič [27, 28, 29].

The zero element of WC is the class of the space $(\mathcal{D},\mu)$ where $\mathcal{D}=\mathcal{C}$ and μ is just the law of addition on $\mathcal{C}$. A homogeneous space $(\mathcal{D},\mu)$ is in the zero class if and only if there is a point on $\mathcal{D}$ defined over k.

This last remark calls attention to a disadvantage in the notion of a class of principal homogeneous spaces. There is no known procedure which will infallibly tell us whether a curve $\mathcal{D}$ defined over k has a point on it defined over k or not. Hence we cannot always decide whether two principal homogeneous spaces are in the same class or not. To some extent it is possible to obviate this disadvantage and work with only effectively computable objects at the expense of some complication in the enunciations. I shall not go into this point further here.

4. A principal homogeneous space defined over a field k is clearly also defined over any field K containing k. Hence there is a natural map

$$WC(\mathcal{C},k)\to WC(\mathcal{C},K), \tag{4.1}$$

which can easily be seen to be a group homomorphism. In particular, k may be an algebraic numberfield and K one of its completions k_v with respect to a valuation v. It turns out that a given element of $WC(\mathcal{C},k)$ is in the kernel of the localisation map

$$j_v\colon \qquad WC(\mathcal{C},k)\to WC(\mathcal{C},k_v) \tag{4.2}$$

for all except a finite number of v. (This is the same as saying that a curve $\mathcal{D}$ of genus 1 defined over k has a point defined over k_v for all except finitely many v). We may thus define groups $Ш$ and $Б$ by the exactness of the sequence

$$0\to Ш\to WC(k)\to\sum_v WC(k_v)\to Б\to 0. \tag{4.3}$$

The Tate-Šafarevič group $Ш$ thus consists of the classes of principal homogeneous spaces $(\mathcal{D},\mu)$ which are "everywhere locally trivial", i.e. such that there is everywhere locally a point on $\mathcal{D}$. So far as I know the first example of a curve $\mathcal{D}$ which has a point everywhere locally but not globally was given by Reichardt [26]. He showed that there are no rational points on

$$x^4-17y^4=2z^2, \tag{4.4}$$

although there certainly are everwhere locally. Later Selmer gave many examples, e.g.

$$3x^3+4y^3+5z^3=0. \tag{4.5}$$

The group $Ш$ is intimately connected with the problem of determining the group $\mathfrak{G}$ of points on $\mathcal{C}$ defined over k. Let m be a positive integer and let Δ_m be the set of points on $\mathcal{C}$ of order dividing m, whether defined over k or not. There is then an exact sequence

$$0\to\Delta_m\to\overline{\mathfrak{G}}\xrightarrow{m}\overline{\mathfrak{G}}\to 0, \tag{4.6}$$

where, as before, $\overline{\mathfrak{G}}$ is the group of points defined over the algebraic closure $\bar{k}$ of k and where $\xrightarrow{m}$ is multiplication by m. Since

$$H^0(\Gamma,\overline{\mathfrak{G}})=\mathfrak{G}, \quad H^1(\Gamma,\overline{\mathfrak{G}})=WC, \tag{4.7}$$

the associated cohomology exact sequence gives us the exact sequence

$$0\to\mathfrak{G}/m\mathfrak{G}\to H^1(\Gamma,\Delta_m)\to(WC)_m\to 0, \tag{4.8}$$

where $(WC)_m$ is the group of elements of WC of order dividing m.

Hence we have an exact sequence

$$0 \to \mathfrak{G}/m\mathfrak{G} \to S^{[m]} \to (Ш)_m \to 0, \tag{4.9}$$

where $S^{[m]}$ is the subgroup of $H^1(\Gamma, \Delta_m)$ consisting of elements which, in an appropriate sense, come from $\mathfrak{G}/m\mathfrak{G}$ "everywhere locally" and where $(Ш)_m$ is the subgroup of $Ш$ of elements of order dividing m. For any given m the group $S^{[m]}$ is finite and can always be effectively determined. When $S^{[m]}$ is known, the exact sequence (4.9) gives us an upper bound for the number of generators of $\mathfrak{G}$: and then $(Ш)_m$ is a measure of the error in this upper bound. (For all this see Lang and Tate [15] or Šafarevič [27, 28, 29]).

In particular we may sometimes be able by good luck to find points on C defined over k which generate a group $\mathfrak{G}_0$ such that

$$\text{order } \mathfrak{G}_0/m\mathfrak{G}_0 = \text{order } S^{[m]}. \tag{4.10}$$

Then $\mathfrak{G}_0$ is of finite index in $\mathfrak{G}$; and $\mathfrak{G}$ can be completely determined using Fermat's "descente infinie" as in the proofs of the Mordell-Weil theorem. All(1) the determinations of $\mathfrak{G}$ in special cases seem to follow this pattern, although the point of view of, say, Fermat was rather different.

5. Because of the exact sequence (4.9) something was known about the Tate-Šafarevič group $Ш$ before it was invented. Selmer [31, 32, 33, 34] in effect(2) compared the estimate g_m for the number of generators of $\mathfrak{G}$ using (4.9) with the corresponding estimate g_{m^2} using m^2 instead of m. He obtained a vast amount of numerical evidence that the difference $g_m - g_{m^2}$ must be even. It is natural to conjecture that an even number must be somehow the rank of a skew-symmetric matrix, and in fact it can be shown(3) that there is a naturally-defined skew-symmetric bilinear form l on $Ш$ taking values in the group $\mathbf{Q}/\mathbf{Z}$ of the rationals modulo 1 [6, 7, 8, 9]. The kernel of l, i.e. the set of $\xi \in Ш$ such that $l\,\xi, \eta) = 0$ for all $\eta \in Ш$, is precisely the group $Ч$ of elements of $Ш$ which can be divided by any positive integer (the "infinitely divisible" elements). Hence l sets the quotient group $Ш/Ч$ in duality with itself.

It is an easy consequence of the finiteness of the group $S^{[m]}$ in (4.9) that the q-primary part of $Ш/Ч$ is finite for each prime q. The numerical work of Selmer gives some reason(4) to suppose that the infinitely divisible part $Ч$ of the Tate-Šafarevič group $Ш$ consists solely of the zero element. Indeed Tate and Šafarevič have, I believe, independently conjectured(5) that $Ш$

(1) Perhaps Reichardt [26] should be cited as an exception. He uses the method described above to determine not $\mathfrak{G}$ but the group $\mathfrak{G}_K$ of points defined over a finite normal extension K of k. (In his case $k = \mathbf{Q}$ is the rationals and $K = \mathbf{Q}(\sqrt{-2})$. Then $\mathfrak{G}$ is the set of elements of $\mathfrak{G}_K$ fixed by the automorphism group of K/k.

(2) Selmer was mainly interested in a case with complex multiplication and his m was usually $\sqrt{-3}$.

(3) It is natural to conjecture that the results of this section go over to abelian varieties of any dimension or at least to all jacobians. There are partial results in Poitou [22, 23, 24]. The results announced by Blanchard and Poitou [23] cannot, however, be regarded as proved (see Poitou [25]). During the Congress Tate showed me how the generalization can be effected. (Cf. his lecture.)

(4) The difference between the estimate g_m and the true number g of generators is always even when the latter can be determined (e.g. by using (4.9) with a number prime to m instead of m).

(5) In his lecture Tate denied paternity but adopted the conjecture. In conversation during the Congress Šafarevič expressed strong doubts.

itself is always finite, although, so far as I know, it has not been completely determined in any individual case.(1)

Under the hypothesis that Ш is finite one can determine the exact sequence of compact groups which is dual to the exact sequence

$$0 \to Ш \to WC(k) \to \Sigma WC(k_v) \to Б \to 0 \tag{5.1}$$

of discrete groups. Tate [37] has shown that the dual of the local group $WC(k_v)$ is

$$\mathfrak{G}'_v = \begin{cases} \mathfrak{G}_v & \text{if } v \text{ is finite} \\ \mathfrak{G}_v & \text{modulo its connected component if } v \text{ is infinite} \end{cases}$$

$$= \varprojlim_n \mathfrak{G}_v / n!\, \mathfrak{G}_v \quad \text{(in both cases).} \tag{5.2}$$

The finite group Ш is its own dual. The dual of Б is the closure $\tilde{\mathfrak{G}}$ of the group $\mathfrak{G}$ of global points in the embedding

$$\mathfrak{G} \to \prod_v \mathfrak{G}'_v. \tag{5.3}$$

It seems probable that(2) $$\tilde{\mathfrak{G}} = \varprojlim_n \mathfrak{G}/n!\, \mathfrak{G} \tag{5.4}$$

with the limit topology.

Finally, the dual, Θ, say, of the global group WC is roughly speaking the group of adeles modulo principal adeles. An adele consists by definition of an element of Ш represented by a homogeneous space $(\mathcal{D}, \mu)$ together with a set of points $\mathfrak{a}_v$ on $\mathcal{D}$ defined over k_v, where v runs through all the valuations of k. The adeles have a natural topology and group law which I need not describe. The principal adeles are those for which $\mathfrak{a}_v$ is a global point independent of v. More precisely, Θ is the quotient group of the adele group for which the sequence

$$0 \leftarrow Ш \leftarrow \Theta \leftarrow \prod_v \mathfrak{G}'_v \leftarrow \tilde{\mathfrak{G}} \leftarrow 0 \tag{5.5}$$

is exact. Then (5.5) is dual(3) to (5.1). The description of Θ is reminiscent both of classfield theory and of recent work on the arithmetic of linear algebraic groups (see Weil [43], Borel [3], Borel's lecture to the Congress, and the discussion later in this talk).

Šafarevič [30] and Ogg [21] have obtained somewhat similar results when the groundfield is a functionfield of transcendence degree 1 over an algebraically closed field. See also the laconic report of Tate [38].

---

(1) Only the q-primary part for a finite number of primes q. Šafarevič told me during the conference that he could show that the q-primary part is infinitely divisible for all except a finite number of q. This implies that $Ш/Ч$ is finite. Borel in his lecture stated that he and Serre had shown that the analogue of Ш for linear algebraic groups is finite.

(2) During the Congress I learned that Serre has proved this (for elliptic curves only, not all abelian varieties).

(3) If the q-primary part of Ш is finite then the q-primary parts of (5.1) and (5.5) are dual even though Ш may not be finite.

6. With their permission I should like now to describe some recent work of Birch and Swinnerton-Dyer [1]. As a result of a great series of carefully and skillfully planned numerical experiments on the Cambridge computer EDSAC they have evolved some very precise and highly plausible conjectures. So far no proofs have been found, but the conjectures fit happily into the framework both of known results and of some widely publicized unproved conjectures.

They take the groundfield to be the rationals and consider curves of the special type

$$y^2 = x^3 - Ax - B, \tag{6.1}$$

where A and B are rational integers, not both zero. We shall say that a rational prime p is good if the equation (6.1) taken modulo p is the equation of a curve of genus 1 over the field of p elements. All except a finite number of p are good. Let N_p be the number of points on the modulo p curve, so N_p is the number of solutions (x,y) of the congruence

$$y^2 = x^3 - Ax - B \pmod{p} \tag{6.2}$$

plus 1 (for the "point at infinity"). It is known that

$$|N_p - (p+1)| < 2\sqrt{p}. \tag{6.3}$$

It is natural to suppose that N_p will tend on the whole to be larger if there are a lot of rational points on the curve, i.e. if the number g of generators of infinite order of the group $\mathfrak{G}$ of rational points is large. Calculations of

$$\prod_{p \leqslant 2000} \frac{N_p}{p} \tag{6.4}$$

for a large number of curves (6.1) produced good confirmation.

We now recall that if p is a good prime the "congruence zeta-function" is given by

$$\zeta_p(s) = \frac{f_p(s)}{(1-p^{-s})(1-p^{1-s})}, \tag{6.5}$$

where
$$f_p(s) = 1 + (N_p - p - 1)\, p^{-s} + p^{1-2s}. \tag{6.6}$$

The product
$$f(s) = \prod_{p \text{ good}} f_p(s) \tag{6.7}$$

converges for $\mathcal{R}s > \frac{3}{2}$. It was conjectured by Hasse (1) that $f(s)$ can be meromorphically continued over the whole plane. We have

$$f_p(1) = N_p/p \tag{6.8}$$

and so, quite formally,
$$f(1) = \prod_{p \text{ good}} \frac{N_p}{p}. \tag{6.9}$$

(1) Hasse's conjecture was expressed in terms of $\Pi_p \zeta_p(s) = f(s)\zeta(s)\zeta(s-1)$ (apart from possible trivial factors coming from the bad primes). The conjecture has since been generalized by Weil.

The work of Birch and Swinnerton-Dyer leads them to conjecture that $f(s)$ has a pole at $s=1$ whose order is the number g of generators of $\mathfrak{G}$ of infinite order.

There is, of course, an analogue of this Birch-Swinnerton-Dyer conjecture in which the groundfield is a field of transcendence degree 1 over a finite field. According to the experts, this analogue fits neatly (1) into the familiar complex of conjectures about varieties over finite fields. Tate tells me (2) that one case of the analogue has been proved by Mumford.

7. In the case
$$y^2 = x^3 - Ax \tag{7.1}$$

Birch and Swinnerton Dyer could carry their investigation much further. As a special instance (3) of a result of Deuring [12] the function

$$\{f(s)\}^{-1} = \prod_{p \text{ good}} \frac{1}{1 + (N_p - p - 1)\, p^{-s} + p^{1-2s}} \tag{7.2}$$

is a Hecke L-function. Birch and Swinnerton-Dyer have found an expression for $\{f(1)\}^{-1}$ as a rather elaborate finite sum from which they deduce that

$$\frac{\tilde{\omega}}{f(1)} = \begin{cases} A^{\frac{1}{2}} i_A & \text{if} \quad A > 0, \\ (-4A)^{\frac{1}{2}} i_A & \text{if} \quad A < 0, \end{cases} \tag{7.3}$$

where $\tilde{\omega}$ is the real halfperiod of the Weierstrass $\wp$-function for which $\wp'^2 = 4(\wp^3 - \wp)$ and i_A is a rational integer (at least if A is not divisible by the fourth power of a prime, which we can suppose without loss of generality). From the formal expression (6.9) for $f(1)$ there is thus a sense in which

$$\left.\begin{array}{c} \dfrac{\tilde{\omega}}{A^{\frac{1}{2}} i_A} \\ \text{or} \quad \dfrac{\tilde{\omega}}{(-4A)^{\frac{1}{2}} i_A} \end{array}\right\} = \prod_{p \text{ good}} \frac{N_p}{p}, \tag{7.4}$$

although we do not know that the product on the right hand side converges in the conventional sense.

The conjecture about the order of the pole of $f(s)$ at $s=1$ suggest, in particular, that $i_A = 0$ when and only when there are infinitely many rational points on (7.1). This is a precise conjecture which has been checked for a large number of values of A.

I should like to emphasize the revolutionary nature of this conjecture. (4) Hitherto the only way of proving the existence of rational points on a curve of genus 1 has been by inspection, i.e. by actually writing down their

(1) If I understand the position correctly, the Birch-Swinnerton-Dyer conjecture is compatible with but not obviously implied by the other conjectures mentioned.

(2) Cf. also Tate's lecture to the Congress.

(3) For (7.1), however, there are explicit formulae for N_p in terms of the representation of p as the sum of two squares which are much older; and the stated result follows from these. Deuring's results have been generalized, see Shimura and Taniyama [35].

(4) Tate pointed out during the Congress that the conjecture that $Ш$ is a finite group is equally revolutionary in view of Šafarevič's unpublished result that the q-primary component of $Ш$ is infinitely divisible for all except a finite computable set of q.

coordinates and showing that they satisfy the relevant equations. Birch and Swinnerton-Dyer conjecture that by computing i_A we can show the existence of points without actually finding them. It is perhaps reasonable to hope that any proof of the conjecture would bring with it a way of finding the points i.e. would give us an effective means of finding the group $\mathfrak{G}$.

When $i_A \neq 0$, the numerical work also produced the totally unexpected conjecture that

$$i_A = \text{(power of 2) (perfect square)}. \tag{7.5}$$

Birch and Swinnerton-Dyer interpret the perfect square as the order of the Tate-Šafarevič group $Ш$. This latter order, if finite, must be a perfect square because of the existence of the skew-symmetric bilinear form l on $Ш$ which was described earlier.

Some confirmation of this conjecture and a convincing interpretation of the power of 2 in (7.5) is given by the work of Tamagawa(1) on linear algebraic groups,(2) which we must first describe briefly. Let $\mathcal{A}$ be a linear algebraic group defined over an algebraic numberfield k. We denote the group of points on $\mathcal{A}$ defined over k by $\mathfrak{H}$ and, similarly $\mathfrak{H}_v$ is the group of points on $\mathcal{A}$ defined over the local completion k_v of k with respect to the valuation v. There is a canonical way of introducing an invariant measure in the groups $\mathfrak{H}_v$. In many cases it can be shown that the product measure on $\prod_v \mathfrak{H}_v$ converges. Further, $\mathfrak{H}$ is a discrete subgroup of $\prod_v \mathfrak{H}_v$. In many cases it can be shown that the induced measure of the quotient group $\prod_v \mathfrak{H}_v/\mathfrak{H}$ is an integer. For example, the fact that this integer is 2 for an othogonal group is equivalent to some deep results of Siegel about quadratic forms.

In the case of an abelian variety the definition of the canonical measure on the local groups still makes sense. It is, however, no longer true that the the global group is injected in the product of the local groups as a discrete, or even as a closed group, so the analogue of the Tamagawa results certainly cannot hold without some modification. The work of Birch and Swinnerton-Dyer suggests what the analogue should be.

We recollect that if the Tate-Šafarevič group $Ш$ is finite there is an exact sequence

$$0 \leftarrow Ш \leftarrow \Theta \leftarrow \prod_{p \text{ including } \infty} \mathfrak{G}'_p \leftarrow \tilde{\mathfrak{G}} \leftarrow 0 \tag{7.6}$$

where the notation is that introduced earlier except that as we shall be using the rational field as the ground field we write $\mathfrak{G}_p$ for $\mathfrak{G}_v$ where v is the valuation associated with the prime p. Now $\mathfrak{G}'_p = \mathfrak{G}_p$ for $p \neq \infty$ and the measure of the whole group $\mathfrak{G}_p$ in the Tamagawa sense turns out to be just N_p/p when p is a good prime. When p is a bad prime the measure of $\mathfrak{G}_p$ turns out always to be a power of 2, at least in the case (7.1) under discussion. On comparing (7.4) and (7.5) we are led to interpret the power of

(1) Kneser independently started work on similar lines and talked about his work and Tamagawa's at the Edinburgh Congress.

(2) Unfortunately this work has never been published. Two mimeographed accounts by Weil [42, 43] are available. See, however, the lectures of Borel, Kneser and Tate at this Congress.

2 in (7.5) as the contribution of the bad primes (including ∞). and this is confirmed by copious numerical evidence.(1) The interpretation of the perfect square in (7.5) as the order of the Tate-Šafarevič group *Ш* is strengthened by the way in which *Ш* occurs in the exact sequence (7.6).

We are thus led to conjecture that there is some way(2) of defining a measure on the group Θ (which is the dual of the group WC) analogous to the Tamagawa measure for which the measure of the whole group is always the same, say 1. This makes sense, of course, not merely for the special curve (7.1) but for all curves of genus 1. This conjecture is so elegant that it must be true. Its truth implies the truth of the various conjectures made on the way.

8. In conclusion I should like to emphasize how much the work described owes its origin to the need to explain laws found empirically from the inspection of numerical data. When a theorem, say the law of quadratic reciprocity, has been established one is apt to forget that it started life as a conjecture based on numerical evidence. Number theory is an experimental science.

REFERENCES

[1]. Birch, B. J., & Swinnerton-Dyer, H. P. F., Notes on elliptic curves. I. *J. reine angew. Math.*, 212 (1963), 7–25, and later papers of the same title to appear.

[2]. Blanchard, A. & Poitou, G., Sur les composantes primaires du groupe des éspaces homogènes principaux localement triviaux relatifs à une meme variété abélienne. *C.R. Acad. Sci. Paris*, 263 (1961), 2309–2310.

[3]. Borel, A., Some properties of adele groups attached to algebraic groups. *Bull. Amer. Math. Soc.*, 67 (1961), 583–585.

[4]. Cassels, J. W. S., A note on the division values of $\wp(u)$. *Proc. Cambridge Philos. Soc.*, 45 (1949), 167–172.

[5]. —— The rational solutions of the diophantine equation $Y^2 = X^3 - D$. *Acta Math.*, 82 (1950), 243–273. (Addendum *84*, 299.)

[6]. —— Arithmetic on curves of genus 1. I. On a conjecture of Selmer. *J. reine angew. Math.*, 202 (1959), 52–99.

[7]. —— Arithmetic on curves of genus 1. II. A general result. *Ibid.*, 203 (1960), 174–208.

[8]. —— Arithmetic on curves of genus 1. III. The Tate-Šafarevič and Selmer groups. *Proc. London Math. Soc.*, (3) 12 (1962), 259–296.

[9]. —— Arithmetic on curves of genus 1. IV. Proof of the Hauptvermutung. *J. reine angew. Math.* 211 (1962), 95–112.

[10]. —— Arithmetic on curves of genus 1. V. Two counter-examples. (To appear in *J. London Math. Soc.*, 1963.)

[11]. Châtelet, F., Méthode galoisienne et courbes de genre un. *Ann. Univ. Lyon, Section A*, 89 (1946), 40–49.

[12]. Deuring, M., Die Zetafunktion einer algebraischer Kurve vom Geschlechte Eins. I, II, III, IV. *Nach. Akad. Wiss. Göttingen* (1953), 85–94; (1955), 13–42; (1956), 37–76; (1957), 55–80.

[13]. Lang, S., *Abelian varieties*. New York and London, 1959.

(1) Of course it is not excluded that the square in (7.5) can be even.

(2) Of course $\tilde{\mathfrak{G}}$ in (7.6) must be brought in somehow, but it is not clear how. Fortunately in the case (7.1) when this group is finite it is always of order 2 except when A is a square; and so it should have the same effect on the hypothetical measure.

[14]. —— Some theorems and conjectures in diophantine equations. *Bull. Amer. Math. Soc.*, 66 (1960), 240–249.

[15]. Lang, S. & Tate, J., Principal homogeneous spaces over abelian varieties. *Amer. J. Math.*, (1958), 659–684.

[16]. Lutz, E., Sur l'equation $y^2 = x^3 - Ax - B$ dans les corps $\mathfrak{p}$-adiques. *J. reine angew. Math.*, 177 (1937)., 237–247.

[17]. Manin, I. N., (И. Н. Манин). О диофантовых уравнениях над функциональными полями. *Доклады Акад. Наук СССР*, 139 (1961), 806–809.

[18]. Mattuck, A., Abelian varieties over p-adic groundfields. *Ann. Math.*, (2) 62 (1955), 92–119.

[19]. Mordell, L. J., On the rational solutions of the indeterminate equations of the third and fourth degrees. *Proc. Cambridge. Philos. Soc.*, 21 (1922), 179–192.

[20]. Nagell, T., Solution de quelque problèmes dans la théorie arithmetique des cubiques planes du premier genre. *Vid. Akad. Skrifter. Oslo*, I (1935) No. 1.

[21]. Ogg, A. P., *Ann. Math.*, 76 (1962), 185–212.

[22]. Poitou, G., Points rationels sur les courbes. *Séminaire Dubreil-Pisot*, 14 (1960). (Not yet promulgated.)

[23]. —— A propos d'une conjecture de Cassels. *C. R. Acad. Sci. Paris*, 252 (1961), 968–970.

[24]. —— Remarques sur la division des fonctions abéliennes. *C. R. Acad. Sci. Paris*, 252 (1961), 648–650.

[25]. —— Rectification de deux notes précédentes. *C. R. Acad. Sci. Paris*, 254 (1962), 206.

[26]. Reichardt, H., Einige im Kleinen überall lösbare, im Grossen unlösbare diophantische Gleichungen. *J. reine angew. Math.*, 184 (1942), 12–18.

[27]. Šafarevič, I. R., (И. Р. Шафаревич), О бирациональной эквивалентности эллиптических кривых. *Доклады Акад. Наук СССР*, 114 (1957), 267–270.

[28]. —— Показатели эллиптических кривых. *Доклады Акад. Наук*, 114 (1957), 714–716.

[29]. —— Группа главных однородных многообразий. *Доклады Акад. Наук*, 124 (1959), 42–43.

[30]. —— Главные однородные пространства, определенные над полем функций. *Труды Мат. Инст. им. В. А. Стеклова*, 64 (1961), 316–346.

[31]. Selmer, E. S., The diophantine equation $ax^3 + by^3 + cz^3 = 0$. *Acta Math.*, 85 (1951), 203–362.

[32]. —— The diophantine equation $ax^3 + by^3 + cz^3 = 0$. Completion of the tables. *Acta Math.*, 92 (1954), 191–197.

[33]. —— A conjecture concerning rational points on cubic curves. *Math. Scand.*, 2 (1954), 49–54.

[34]. —— The diophantine equation $\eta^2 = \zeta^2 - D$. A note on Cassels' method. *Math. Scand.*, 3 (1956), 68–74.

[35]. Shimura, G. & Taniyama, Y., Complex multipliation of abelian varieties. *Pub. Math. Soc. Japan*, No. 6 (1961).

[36]. Skolem, Th., Diophantische Gleichungen. *Ergebnisse d. Math*, 5, 4 (1938).

[37]. Tate, J., WC-groups over $\mathfrak{p}$-adic fields. *Séminaire Bourbaki*, Exp. 156 (1957).

[38]. —— Principal homogeneous spaces for abelian varieties. *J. reine angew. Math.*, 209 (1962), 98–99.

[39]. WEIL, A., L'arithmétique sur les courbes algébriques. *Acta Math.*, 52 (1928), 281–315.

[40]. —— Sur les fonctions elliptiques $\mathfrak{p}$-adiques. *C. R. Acad. Sci. Paris*, 203 (1936), 22–24.

[41]. —— On algebraic groups and homogeneous spaces. *Amer. J. Math.*, 77 (1955), 493–512.

[42]. —— Adèles et groupes algébriques. *Séminaire Bourbaki*, Exp. 186 (1958).

[43]. —— *Adeles and Algebraic Groups*. (Lecture notes.) Institute for Advanced Study, Princeton, 1961.

(Added in proof):

[44]. CHABAUTY, C., Sur les points rationnels des courbes algébriques de genre supérieur à l'unité. *C. R. Acad. Sci. Paris*, 212 (1941), 882–885.

[45]. MAHLER, K., On the division values of Weierstrass' $\boldsymbol{p}$-function. *Quart. J. Math.*, (2) 6 (1935), 76–77.

A DEFORMATION THEORY FOR THE ZETA FUNCTION OF A HYPERSURFACE(1)

By BERNARD DWORK

Introduction

Let $\mathfrak{H}$ be a non-singular hypersurface of degree d in projective n-space of characteristic p defined over the field of q elements. Let N_s be (for each integer $s \geqslant 1$) the number of points of $\mathfrak{H}$ which are rational over the field of q^s elements. The zeta function of $\mathfrak{H}$ is defined [1] by

$$\zeta((\mathfrak{H}, t) = \exp\left\{\sum_{s=1}^{\infty} N_s t^s/s\right\}. \tag{1}$$

The conjectures of Weil [1] concerning the zeta functions of non-singular varieties imply (or suggest) the following three conjectures for hypersurfaces.

I.

$$\zeta(\mathfrak{H}, t) = P(t)^{(-1)^n} / \prod_{i=1}^{n-1} (1 - q^i t), \tag{2}$$

where P is a polynomial of degree $N = d^{-1}\{(d-1)^{n+1} + (-1)^{n+1}(d-1)\}$.

The reciprocals $\omega_1, \omega_2, \ldots, \omega_n$ of the zeros of P are conjectured to satisfy the conditions

II. (Functional Equation)

$$\omega_i \to q^{n-1}\omega_i \quad \text{is a permutation of the} \quad \omega_i.$$

III. (Riemann Hypothesis)

$$|\omega_i| = q^{(n-1)/2} \quad (i = 1, 2, \ldots, N),$$

i.e. as complex numbers the ω_i have magnitude independent of i.

We have verified conjecture I provided $1 = (2, p, d)$, i.e. provided either p or d is odd [3]. The remaining conjectures remain open but as is well known the conjectures of Weil have been verified for varieties of dimension 1, for abelian varieties [2], for hypersurfaces defined by equations of type $\sum_{i=1}^{n+1} a_i X_i^d$, for rational varieties of dimension 2, for algebraic homogenous spaces and for certain varieties constructed from varieties of the preceding types.

After defining a number of symbols, we shall continue our introductory remarks.

Q' = field of rational p-adic numbers.

Ω = completion of algebraic closure of Q'.

(1) This work was partly supported by U.S. Army, Office of Ordinance Research Grant DA-ORD-31-124-61-G95 and by an Alfred P. Sloan Foundation Grant.

$\mathfrak{O}$ = ring of integers of Ω = "closed" unit disk in Ω.
$\mathfrak{P}$ = ideal of non units in $\mathfrak{O}$ = "open" unit disk in Ω.
ord denotes the (additive) valuation of Ω, normalized by ord $p=1$.
Z = ring of rational integers.
Z_+ = set of all non-negative elements of Z.
S = set $\{1,2,\ldots,n+1\}$.
$\mathfrak{H}_A$ = intersection of $\mathfrak{H}$ with the subspace $\{X_i=0\}_{i\notin A}$, A being a subset of S.
If G is a square matrix, G^t denotes its transpose.
$\mathfrak{T}=\{u=(u_0,u_1,\ldots u_{n+1})\in Z_+^{n+1}\,|\,du_0=u_1+\ldots+u_{n+1}\}$.
Teichmüller points $=\{x\in\Omega\,|\,x^{q^r}=x$ for some $r\in Z\}$.
$\mathcal{A}=\{u\in\mathfrak{T}\,|\,0\leqslant u_i<d,\ i=1,2,\ldots,n+1\}$.
$\mathcal{A}'=\{u\in\mathcal{A}\,|\,u_i>0,\ i=1,2\ldots,n+1\}$.
π = fixed element in Ω such that $\pi^{p-1}=-p$.
ψ = endomorphisms of $\Omega\{X_0,X_1,\ldots,X_{n+1}\}$, as vector space over Ω, defined by linearity over Ω and $\psi(X^u)=\begin{cases}X^{u/q} & \text{if } q\,|\,u,\\ 0 & \text{otherwise,}\end{cases}$ for each $u\in Z_+^{n+2}$.
Φ = endomorphism of $\Omega\{X_0,\ldots,X_{n+1}\}$ (or of $\Omega\{1/X_0,\ 1/X_1,\ldots,1/X_{n+1}\}$) defined by $X^u\to X^{qu}$.
δ is the endomorphism $g\to g(t)^\delta=g(t)/g(qt)$ of $1+t\Omega\{t\}$.
γ_- is endomorphism of space of all sums $\sum_{u\in Z^{n+2}} B_uX^u$ defined by

$$\gamma_-(X^u)=\begin{cases}X^u & \text{if } u_i\leqslant 0,\ i=0,1,\ldots,n+1,\\ 0 & \text{otherwise.}\end{cases}$$

$$(r)_m=\begin{cases}1 & \text{if } m=0\\ r(r+1)\ldots(r+m-1) & \text{if } m\in Z,\ m\geqslant 1.\end{cases}$$

A homogeneous polynomial $\bar{f}(X_1,X_2,\ldots,X_{n+1})$ in $n+1$ variables with coefficients in the residue class field of Ω is said to be *regular* if the polynomials $\bar{f}$, $X_1(\partial\bar{f}_1/\partial X_1)$, $X_2(\partial\bar{f}_2/\partial X_2)$, ..., $X_{n+1}(\partial\bar{f}_{n+1}/\partial X_{n+1})$ have no common zero in projective n-space.

A homogeneous polynomial, $f(X_1,\ldots,X_{n+1})$ with coefficients in $\mathfrak{O}$ will be said to be *regular* if its image $\bar{f}$ under reduction modulo $\mathfrak{P}$ is regular.

It is easily seen that the irreducible defining polynomial of a given hypersurface $\mathfrak{H}$ is regular if and only if $\mathfrak{H}_A$ is non-singular for each subset A of S. We suppose that $\mathfrak{H}$ is given with a specified inbedding in projective space and hence it will make sense to say that $\mathfrak{H}$ is *regular* if its irreducible defining polynomial is regular.

We have shown [3, § 4c] that very little is lost by restricting our attention to regular hypersurfaces. The hypothesis of regularity will often be used, but when used it will be explicitly indicated.

The central point in our previous article is that if $\mathfrak{H}$ is regular and $1=(2,p,d)$ then $P(qt)$ is the characteristic polynomial an endomorphism, β (denoted $\bar{\alpha}_S^S$ in [3]) of a certain finite dimensional vector space, $\mathfrak{W}^S$ (denoted $\mathfrak{W}_S^S$ in [3]). We remark that very little is known about this endomorphism, for example we do not even know if it is semi-simple. We shall take this opportunity (§ 6) to give some estimates for the p-adic valuation of the coefficients of a matrix representing β and from this will deduce estimates for the Newton polygen of $P(t)$, i.e. deduce estimates for the p-adic values of the zeros of P. These estimates are merely consequences of the functional equation in the case of plane curves but for hypersurfaces of higher

dimensions these estimates are consistant with, but not covered, by the conjectures of Weil.

A brief account of the construction of β (for $\neq 2$) is given § 1. In § 2 we construct an endomorphism, β^*, dual to β and in § 3 we examine this construction explicitly in the case of "diagonal forms" of type $\sum_{i=1}^{n+1} a_i X_i^d$. The main object of this report (§ 4) is to give a method by which diagonal forms may be deformed into more general homogeneous forms in such a way that results valid for diagonal forms can be carried over to the general case. We restrict our attention (in § 4) to the case $p \nmid d$, $p \neq 2$ and consider a one parameter family of homogeneous polynomials, f_Γ, of degree d in $X_1, X_2, \ldots, X_{n+1}$ with coefficients in $\mathfrak{O}$. Let W be the set of all $\Gamma \in \mathfrak{O}$ such that f_Γ is regular. For each $\Gamma \in W$ we associate with Γ (in a manner suggested by § 2) a mapping β_Γ^* between two N dimensional vector spaces. When $\Gamma^q = \Gamma$, β_Γ^* is an endomorphism whose characteristic polynomial determines the zeta function of the hypersurface $f_\Gamma(X) \equiv 0 \bmod \mathfrak{P}$. The mapping β_Γ^* is associated with a matrix, (β_Γ^*), which is meromorphic (in the sense of Krasner [7]) in W. We now construct an $N \times N$ matrix, C_Γ, whose coefficients are power series in Γ with coefficients which are rational functions of the coefficients of f_Γ. We find that for ord $\Gamma > 0$,

$$(\beta_\Gamma^*) = C_{\Gamma^q}^{-1}(\beta_0^*)\, C_\Gamma, \tag{3}$$

(β_0^*) being the matrix obtained from (β_Γ^*) by specializing $\Gamma = 0$. The matrix C_Γ is the matrix of solutions of a set of N linear ordinary differential equations whose coefficients are rational functions of Γ. Furthermore C_Γ is formally independent of p and hence we are led to the conjecture that the matrix has a classical interpretation. The theory of Krasner shows that equation (3) specifies (β_Γ^*) on W.

In § 4 we use an $N \times N$ matrix, J, (constructed in § 3) independent of p, whose non-zero coefficients are ± 1 and show (again using the uniqueness theorem of Krasner) that the functional equation certainly holds for hypersurfaces obtained by reduction modulo $\mathfrak{P}$ from f_Γ (for $\Gamma \in W$) provided the matrix

$$M_\Gamma = C_\Gamma^t J C_\Gamma$$

is meromorphic in W. We have examined this matrix for the families

(i) $X_1^2 - 2\Gamma X_1 X_2 + X_2^2$,

(ii) $X_1^3 + X_2^3 + X_3^3 - 3\Gamma X_1 X_2 X_3$,

(iii) $X_1^4 + X_2^4 + X_3^4 - 4\Gamma X_1^2 X_2 X_3$,

(iv) $X_1^4 + X_2^4 + X_3^4 + X_4^4 - 4\Gamma X_1 X_2 X_3 X_4$,

and can report that in each case the coefficients of M_Γ are rational functions of Γ. In particular this shows that the functional equation is valid for all non-singular fourth degree surfaces in three space which are represented by a polynomial in family (iv).

We call attention to the fact that the theory may be applied to hypersurfaces of dimension zero, i.e. to one parameter families of homogeneous polynomials of degree d in two variables. In this case equation (3) may be viewed as a generalization of the well known fact that, if $q \equiv 1 \bmod d$, the

splitting of the polynomial $X^d - \Gamma Y^d$ over the field of q elements is determined by the value of $\Gamma^{(q-1)/d}$.

In § 5 we discuss the family of elliptic curves in the Legendre normal form. These polynomials are not regular and the theory of § 4 does not appear to be applicable, but using unpublished results of Tate and Lazard we give indications of the existence of a deformation theory, involving a close connection between hypergeometric series and the zeros of the zeta functions of elliptic curves. We became aware of this connection in 1958; it was the first suggestion of a connection between p-adic analysis and the theory of zeta functions.

§ 1. The endomorphism, β

In this section we define the endomorphism, β, referred to in the introduction. The discussion is essentially the same as in [3] (where the endomorphism was denoted $\bar{\alpha}_S^S$). We shall subsequently point out some differences which are of importance. For b rational, let $L(b)$ be the subspace of the space of power series $\Omega\{X_0, X_1, \ldots, X_{n+1}\}$ in $n+2$ variables defined by

$$L(b) = \left\{ \sum_{u \in \mathfrak{X}} B_u X^u \,\middle|\, \operatorname{ord} B_u - b u_0 > -\infty \right\}.$$

We shall use this symbol only with $b > 0$. Let f be a homogeneous polynomial in $X_1, \ldots, X_{n+1}$ with coefficients in $\mathfrak{O}$ whose reduction, $\bar{f}$, modulo $\mathfrak{P}$ is the defining polynomial of $\mathfrak{H}$. We shall assume that the coefficients of f are sufficiently close to the Teichmüller representatives in $\mathfrak{O}$ of the coefficients of $\bar{f}$. More precisely we impose the condition,

$$\operatorname{ord}\,(c^q - c) > \operatorname{Max}\left(\frac{p-1}{p}, \frac{1}{p-1}\right) \tag{4}$$

for each coefficient, c, of f.

Let $\hat{F}(X) = \exp\,(\pi X_0 f(X)) \in \Omega\{X\}$. We have shown that $F(X) = \hat{F}(X)/\hat{F}(X^q) \in L((p-1)/pq)$ and hence for $(p-1)/p \geqslant b$ we seen that $\xi \to \alpha\xi = \psi(F\xi) = ((1/\hat{F}) \circ \psi \circ \hat{F})\xi$ is an endomorphism of $L(b)$. If (α) denotes the infinite matrix representing α relative to the monomial basis of $L(b)$, we know that det $(\mathrm{I} - t(\alpha))$ is an entire function of t in Ω and we have shown without any hypothesis of non-singularity that

$$\det\,(I - t(\alpha))^{\delta^{n+1}} = (1-t) \prod_A P_A(qt), \tag{5}$$

the product being over all subsets, A, of $S = \{1, 2, \ldots, n+1\}$ where for each A, P_A is the rational function defined by

$$\zeta(\mathfrak{H}_A, t) = P_A\,(t)^{(-1)^{m_A}} \Big/ \prod_{i=0}^{m_A - 1} (1 - q^i t),$$

$1 + m_A$ being the number of elements in A.

To remove the operator δ^{n+1} in equation (5) we introduce differential operators $D_0, D_1, \ldots, D_{n+1}$ in $\Omega\{X\}$ defined by

$$D_i = \frac{1}{\hat{F}(X)} \circ E_i \circ \hat{F}(X),$$

where $E_i = X_i \; \partial/\partial X_i$, $i = 0, 1, \ldots, n+1$. These commute with α according to the rule

$$qD_i \circ \alpha = \alpha \circ D_i$$

and if f is regular then $\mathfrak{W} = L(b)/\sum_{i=1}^{n+1} D_i L(b)$ is, for $p/(p-1) \geqslant b > 1/(p-1)$, a vector space of dimension d^n and by passage to quotients we obtain (for $p \neq 2$) a non-singular endomorphism, $\bar{\alpha}$, of $\mathfrak{W}$ with the property

$$\det\,(\mathrm{I} - t\bar{\alpha}) = \det\,(\mathrm{I} - t(\alpha))^{\delta^{n+1}}. \tag{6}$$

Finally, let $\mathfrak{W}^S$ be the subspace of $\mathfrak{W}$ consisting of the image in $\mathfrak{W}$ of $X_1 X_2 \ldots X_{n+1} L(b)$ and let β be the restriction of $\bar{\alpha}$ to $\mathfrak{W}^S$. We now have

$$\det\,(I - t\beta) = P_S(qt) = P(qt). \tag{7}$$

The proof of conjecture I (for $p \neq 2$) follows from this.

Note. The above account differs from [3] in the following respects:

(1) By replacing $\hat{F}$ by a more complicated power series, equation (6) was proven for all p and equation (7) for $1 = (2, p, d)$.

(2) In [3, § 3], where we gave an account of the differential operator theory, we replaced (in our present notation) $L(b)$ by $L(b) \cap \Omega_0\{X\}$, where Ω_0 is a finite extension of Q'. Furthermore the differential operators were assumed to be "defined" over Ω_0. We can now report that all the results (with one exception, namely in Lemmas 3.6 and 3.15 we exclude $b = 1/(p-1)$) of that section remain valid in the general case, that new proofs are needed only for Lemma 3.2 and Lemma 3.10, that the new proof of Lemm 3.2 is a relatively minor modification of the previous one and that the new proof of Lemma 3.10 is precisely the same as the given proof of Lemma 3.9 rewritten so as to apply to the case in which $\varrho = +\infty$.

(3) Our treatment of the spectral theory [3, § 2] was qualified at several points by a similar replacement of $L(b)$ by $L(b) \cap \Omega_0\{X\}$ and by a similar restriction on the operator α. Everything needed here can be obtained by the arguments used there, but a very general treatment of the spectral theory has been made available by Serre [5].

(4) In our previous work [cf. eqn. (4.16), ref. 3] the polynomial f was chosen such that each coefficient c satisfied the condition $c^q = c$. This condition leads to a simplification of the formal proofs, but the Corollary of Theorem 2.5 in [3] shows that condition (4) above is adequate.

§ 2. Dual theory

Our dual theory is necessarily of a formal type since we have not assigned a topological structure to $L(b)$.

Let L_- be the subspace of $\Omega\{1/X_0, 1/X_1, \ldots, 1/X_{n+1}\}$ consisting of all sums

$$\left\{\sum_{v\in\mathfrak{T}} B_v \frac{1}{X^v}\,\middle|\, B_v \in \Omega\right\}.$$

For b' rational, let

$$L_-(b') = \left\{\sum_{u\in\mathfrak{T}} B_u \frac{1}{X^u}\,\middle|\, \operatorname{ord} B_u + b' u_0 > -\infty\right\}.$$

For $b > b'$ we define a pairing of $L(b)$ and $L_-(b')$ into Ω. For $\xi^* \in L_-(b')$, $\xi \in L(b)$ we define $\langle \xi^*, \xi\rangle$ to be the constant term of the formal product $\xi^*(-X_0, X_1, X_2, \ldots, X_{n+1}) \cdot \xi(X_0, X_1, \ldots, X_{n+1})$.

In the following we shall suppose $p \neq 2$, $p/(p-1) \geqslant b > 1/(p-1)$, $b > b' > 0$. Relative to the pairing, the endomorphism α of $L(b)$ has a dual map

$$\alpha^* = \gamma_- \circ \frac{1}{F(X)} \circ \Phi.$$

We also obtain duals $D_0^*, D_1^*, \ldots, D_{n+1}^*$ of the differential operators; explicitly: for $\xi^* \in L_-(b')$, $-D_i^* \xi^* = E_i \xi^* + \pi\gamma_-(\xi^* E_i(X_0 f))$, $i = 0, 1, \ldots, n+1$, an explicit formula which gives an extension of D_i^* to L_-. Let $\mathfrak{K}^*$ be the set of all solutions in L_- of the simultaneous equations

$$D_i^* \xi^* = 0 \quad (i = 1, 2, \ldots, n+1).$$

It can be shown that $\dim \mathfrak{K}^* = d^n$, that $\mathfrak{K}^*$ lies in $L_-(b')$ for each $b' > 0$ and clearly the given pairing of $L_-(b')$ with $L(b)$ induces by restriction a pairing of $\mathfrak{K}^*$ with $L(b)$ and by passage to quotients a pairing of $\mathfrak{K}^*$ with $\mathfrak{W}$. Since $\mathfrak{K}^*$ is invariant under α^*, we obtain, by restriction, an endomorphism $\bar{\alpha}^*$ of $\mathfrak{K}^*$ which is the dual of the endomorphism $\bar{\alpha}$ of $\mathfrak{W}$.

It is natural to consider the annihilator, $\mathfrak{K}^{*(S)}$, of the subspace $\mathfrak{W}^{(S)}$ of $\mathfrak{W}$. Explicitly if $\xi^* = \sum_{v\in\mathfrak{T}} B_v(1/X^v) \in \mathfrak{K}^*$ then $\xi^* \in \mathfrak{K}^{*(S)}$ if and only if $B_v = 0$ for all $v \in \mathfrak{T}$ such that $v_1, v_2, \ldots, v_{n+1}$ are all strictly positive. The factor space $\mathfrak{K}^*/\mathfrak{K}^{*(S)}$ is dual to $\mathfrak{W}^{(S)}$ and we obtain from $\bar{\alpha}^*$ an endomorphism, β^*, of $\mathfrak{K}^*/\mathfrak{K}^{*(S)}$ dual to the endomorphism, β, of $\mathfrak{W}^{(S)}$.

We have noted that $\mathfrak{K}^*$ lies in $L_-(b')$ for each $b' > 0$. A stronger result can also be proven using the fact that $\bar{\alpha}^*$ is a non-singular endomorphism: there exists a constant ϱ such that if $\xi^* = \sum_{u\in\mathfrak{T}} B_u(1/X^u) \in \mathfrak{K}^*$ then $\operatorname{ord} B_u + \varrho \log u_0 > -\infty$.

§ 3. Diagonal forms

Let $f(X) = a_1 X_1^d + \ldots + a_{n+1} X_{n+1}^d$, $p \nmid d$, $p \neq 2$, where the a_i are units in Ω satisfying condition (4).

$$\text{Let } \mathcal{A} = \{u \in \mathfrak{T} \mid 0 \leqslant u_i < d \quad \text{for} \quad i = 1, 2, \ldots, n+1\},$$
$$\mathcal{A}' = \{u \in \mathcal{A} \mid 0 < u_i,\ i = 1, 2, \ldots, n+1\},$$

and let $b = (p-1)/p$.

It is easily seen that $\{\pi^{u_0}X^u\}_{u\in\mathfrak{A}}$ is (in the obvious sense) a basis of $\mathfrak{W}=L(b)/\sum_{i=1}^{n+1} D_iL(b)$ and the complementary basis of $\mathfrak{K}^*$ is $\{\xi^*_{u,0}\}_{u\in\mathcal{A}}$ where for $u\in\mathcal{A}$

$$\xi^*_{u,0}=(-1)^{u_0}\frac{1}{\pi^{u_0}X^u}\sum\prod_{i=1}^{n+1}\left[\left(\frac{u_i}{d}\right)_{j_i}\left(\frac{1}{\pi X_0 a_i X_i^d}\right)^{j_i}\right],$$

the sum being over all $(j_1, \ldots, j_{n+1})\in Z_+^{n+1}$ and the symbol $(r)_m$ appearing in the above formula is defined in the list of symbols.

Likewise $\{\pi^{u_0}X^u\}_{u\in\mathcal{A}'}$ is a basis (in the same sense as above) of $\mathfrak{W}^{(S)}$ and the dual basis of $\mathfrak{K}^*/\mathfrak{K}^{*(S)}$ is the image in that space of the elements $\{\xi^*_{u,0}\}_{u\in\mathcal{A}'}$. A basis of $\mathfrak{K}^{*(S)}$ is $\{\xi^*_{u,0}\}_{u\in\mathcal{A}-\mathcal{A}'}$

For $\xi\in L(b)$ the differential operator, D_i, may be written explicitly as $D_i\xi=E_i\xi+\pi\xi E_i(X_0 f)$, a formula which clearly may be used to extend D_i to the space of all formal sums of type $\sum_{u\in Z^{n+2}} B_u X^u$, where the coeffi cients B_u lie in Ω and there is no restriction on the signs of the components u_i of u (as compared to L_- and to $L(b)$). Since $D_i\circ D_j=D_j\circ D_i$ we may denote by $\prod_{i=1}^{n+1} D_i$ the mapping $D_1\circ D_2\circ\ldots\circ D_{n+1}$ of L_- into this space of formal sums. However for $u\in\mathcal{A}$, we may compute

$$\left(\prod_{i=1}^{n+1} D_i\right)\xi^*_{u,0}=(-1)^{u_0}(\pi d^{n+1}X_0^{n+1}\left[\prod_{i=1}^{n+1}(a_iX_i)^d\right]/(\pi^{u_0}X^u),$$

which shows that the image of $\mathfrak{K}^*$ under this mapping lies in the ring of polynomials $\Omega[X]$ and hence in $L'b)$, while the image of $\mathfrak{K}^{*(S)}$ lies in $\sum_{i=1}^{n+1} D_iL(b)$. By passage to quotients we thus obtain from $(\prod_{i=1}^{n+1}D_i)$ a mapping θ of $\mathfrak{K}^*/\mathfrak{K}^{*(S)}$ onto $\mathfrak{W}^{(S)}$.

Furthermore we can show that

$$\theta\circ\beta^*=q^{n+1}\beta^{-1}\circ\theta,$$

i.e. aside from the factor q^{n+1} the diagram

$$\begin{array}{ccc} K^*/K^{*(S)} & \xrightarrow{\beta^*} & K^*/K^{*(S)} \\ \theta\downarrow & & \downarrow\theta \\ \mathfrak{W}^{(S)} & \xrightarrow[\beta^{-1}]{} & \mathfrak{W}^{(S)} \end{array}$$

is commutative and it is easy to see that this provides a new proof of the known fact [1] that the functional equation holds for the zeta function of hypersurfaces defined by diagonal forms.

We define J to be the matrix of the mapping, $\prod_{i=1}^{n+1}(da_i)^{-1}\theta$ relative to the chosen bases of $\mathfrak{K}^*/\mathfrak{K}^{*(S)}$ and $\mathfrak{W}^{(S)}$. Explicitly, for $u\in\mathcal{A}'$ we set

$$\theta(\xi_{u,0} \bmod \mathfrak{K}^{*(S)})\equiv\sum_{v\in\mathcal{A}'} J_{u,v}X^v \bmod \sum_{i=1}^{n+1} D_iL(b).$$

Let $u\to\bar{u}$ be the permutation of $\mathcal{A}'$ defined by

$$X^{\bar{u}}=\left[\prod_{i=1}^{n+1}(X_0X_i^d)\right]/X^u.$$

It is easy to see that $J_{u,v}=(-1)^{u_0}\delta_{\bar{u},v}$. The matrix J is an $N\times N$ matrix satisfying the conditions $JJ^t=I$

$$J^2=(-1)^{n+1}I,$$

i.e. J is either symmetric or skewsymmetric depending on the parity of n.

§ 4. A one parameter family of polynomials

To fix ideas let $h(X)$ be a homogeneous polynomial of degree d in $n+1$ variables with coefficients in $\mathfrak{O}$ satisfying condition (4). We further assume that the coefficient of X_i^d in h is zero for $i=1,2,\ldots,n+1$. Let $a_1,a_2,\ldots,a_{n+1}$ be units in Ω which also satisfy condition (4). Let Γ be a parameter and we consider the one parameter family

$$f_\Gamma(X)=\sum_{i=1}^{n+1}a_iX_i^d+\Gamma h(X)\qquad (p\nmid d,\ p\neq 2).$$

Let $W=\{\Gamma\in\mathfrak{O}\,|\,f_\Gamma \text{ is regular}\}$. The condition for regularity is easily expressed in terms of the resultant $R(\Gamma)$ of the polynomials $(E_1f_\Gamma,E_2f_\Gamma,\ldots,E_{n+1}f_\Gamma)$. We normalize the resultant by the condition $R(0)=1$. Clearly $W=\{\Gamma\in\mathfrak{O}\,|\,\operatorname{ord}R(\Gamma)=0\}$, which shows that, in the terminology of Krasner, W is a quasi-connected set. We may remark at this point that we restrict our attention to one parameter families because no theory of analytic continuation involving several variables is available.

Let $L_\Gamma=\{\sum_{u\in\mathfrak{T}}B_u(1/X^u)\,|\,B_u\in\Omega(\Gamma)\}$, i.e. L_Γ is similar to L_- in § 2 except that we now allow the coefficients to be rational functions of Γ. We define differential operators $D_{1,\Gamma}^*,D_{2,\Gamma}^*,\ldots,D_{n+1,\Gamma}^*$ precisely as in § 2 except that f is replaced by f_Γ. These differential operators are clearly endomorphisms of L_Γ and we define

$$\mathfrak{K}_\Gamma^*=\{\xi^*\in L_\Gamma\,|\,D_{i,\Gamma}^*\xi^*=0\quad (i=1,2,\ldots,n+1)\}.$$

A pairing of L_Γ and $\Omega(\Gamma)[X]$, the ring of polynomials in $X_0,X_1,\ldots,X_{n+1}$ with coefficients in $\Omega(\Gamma)$, may be defined precisely as in § 2. It takes its values in $\Omega(\Gamma)$ but will be denoted by the same symbol.

We can now state the basic (and not difficult) result of the deformation theory.

THEOREM. *The space K_Γ^* has a basis $\{\xi_{u,\Gamma}^*\}$ indexed by $\mathcal{A}$ dual to $\{\pi^{u_0}X^u\}_{u\in\mathcal{A}}$ and there exists a polynomial $g(\Gamma)\in 1+\Gamma\mathfrak{O}[\Gamma]$ such that for each $u\in\mathcal{A}$*

$$g(\Gamma)\xi_{u,\Gamma}^*=\sum_{w\in\mathfrak{T}}\frac{1}{\pi^{w_0}X^w}G_{w,u}(\Gamma)/R(\Gamma)^{\mu_{w,u}}.$$

where for each $w\in\mathfrak{T}$, $u\in\mathcal{A}$, $\quad G_{w,u}\in\mathfrak{O}[\Gamma]$,

$$\mu_{w,u}\in Z.$$

If Γ is specialized to any element of W, other than a zero of g, then the indicated basis is specialized to a basis of the solutions (in L_-) of the (specialized) equations.

$$D^*_{i,\Gamma}\xi^* = 0 \quad (i = 1, 2, \ldots, n+1).$$

If Γ is specialized to a zero of g in W then the specialized "differential" equations have a solution space (again denoted by $\mathfrak{K}^*_\Gamma$) of dimension d^n, but the elements $\xi^*_{u,\Gamma}$ do not specialize to a basis of $\mathfrak{K}^*_\Gamma$.

If it is desired to specialize Γ to a zero of g then a generic basis capable of specialization can be obtained by changing the set $\mathcal{A}$. However for applications the only specializations of interest are Teichmüller points in W. There exists an element η as close to unity as you please such that none of the zeros of $g(\eta\Gamma)$ are Teichmüller points and hence (changing the coefficients $a_1, \ldots, a_{n+1}$ by the multiplicative factor, η) we may suppose that none of the zeros of g are Teichmüller points in W.

[*Note.* The factor $1/\pi^{w_0}$ in the explicit formula for $\xi^*_{u,\Gamma}$ in the previous theorem may be conservative. We can show that if Γ is specialized to any point in W then $\mathfrak{K}^*_\Gamma$ is specialized to a subspace of $L_-(\varrho)$ for each $\varrho > 0$. If Γ is specialized to a Teichmüller point then the elements of $\mathfrak{K}^*_\Gamma$ satisfy an even stronger growth condition on the coefficients as noted in § 2].

Let $F(X, \Gamma) = \exp\,(\pi X_0 f_\Gamma(X) - \pi X_0^q f_{\Gamma^q}(X^q))$, which we write in the form

$$1/F(X, \Gamma) = \sum_{w \in \mathfrak{T}} A_w(\Gamma) X^w,$$

where for each $w \in \mathfrak{T}$, $A_w(\Gamma)$ is an element of $\Omega[\Gamma]$, which in the obvious valuation of polynomials satisfies the condition: ord $A_w(\Gamma) \geqslant ([p-1]/pq)w_0$.

It is now convenient to think of Γ as specialized to some element of W and hence $\mathfrak{K}^*_\Gamma$ may be viewed as inbedded in $L_-(1/(p-1))$ (and in fact a stronger inbedding statement may be made as noted above). If $g(\Gamma) \neq 0$ then the theorem gives a basis of $\mathfrak{K}^*_\Gamma$.

We define the mapping

$$\alpha^*_\Gamma = \gamma_- \circ \frac{1}{F(X, \Gamma)} \circ \Phi$$

of $\mathfrak{K}^*_{\Gamma^q}$ onto $\mathfrak{K}^*_\Gamma$. The matrix (α^*_Γ) of this mapping relative to the chosen bases of these spaces is given explicitly (if $g(\Gamma)g(\Gamma^q) \neq 0$) by

$$g(\Gamma^q)\,(\alpha^*_\Gamma)_{u,v} = (-1)^{v_0} \sum_{w \in \mathfrak{T}} A_{qw-v}(\Gamma) \frac{1}{\pi^{w_0}} \frac{G_{w,u}(\Gamma^q)}{R(\Gamma^q)^{\mu_{w,u}}} \tag{8}$$

which shows that the coefficients of the matrix (α^*_Γ) are meromorphic in W with poles only at the zeros of $g(\Gamma^q)$. The matrix (α^*_Γ) is invertible if $g(\Gamma) \neq 0$.

We define the supspace $\mathfrak{K}^{*(S)}_\Gamma$ of $\mathfrak{K}^*_\Gamma$ precisely as in § 2 and β^*_Γ is defined as the mapping of $\mathfrak{K}^*_{\Gamma^q}/\mathfrak{K}^{*(S)}_{\Gamma^q}$ into $\mathfrak{K}^*_\Gamma/\mathfrak{K}^{*(S)}_\Gamma$ obtained from α^*_Γ by passage to quotients and the matrix (β^*_Γ) corresponding to the obvious bases is obtained from the matrix (α^*_Γ) by deleting all rows and columns indexed by $u \in \mathcal{A} - \mathcal{A}'$. Clearly the matrix (β^*_Γ) is also meromorphic in W.

If Γ is specialized to an element of W satisfying $\Gamma^q = \Gamma$, then α^*_Γ is an endomorphism of $\mathfrak{K}^*_\Gamma$, and the specialized matrix (β^*_Γ) is the matrix of β^*_Γ relative to the specialized basis and det $(I - t(\beta^*_\Gamma))$ is $P_\Gamma(qt)$, where P_Γ is the polynomial appearing in (2) when $\mathfrak{H}$ is replaced by $\mathfrak{H}_\Gamma$, the

hypersurface defined by reducing the specialized polynomial f_Γ mod $\mathfrak{P}$. Furthermore if Γ were specialized so that $\Gamma^{q^r}=\Gamma$, then $\mathfrak{H}_\Gamma$ is defined over the field of q^r elements and when viewed in that way its zeta function is determined by the characteristic polynomial of the specialization of the matrix $(\beta^{(r-1)})(\beta^{(r-2)})\dots(\beta^{(0)})$, where for typographical reasons, $(\beta^{(j)})$ is written to denote the matrix (β_Γ) with Γ replaced by Γ^{q^j}.

Krasner's theorem of uniqueness of analytic continuation on a quasi-connected domain shows that the matrix (β_Γ) is essentially determined on W if it is specified say in a neighborhood of the origin. We give explicit formulae for β_Γ valid in the "open" unit disk, $\{\Gamma \mid \operatorname{ord}\Gamma>0\}$. Let $G_\Gamma=\exp(\pi\Gamma X_0 h(x))$. For $\Gamma\in\mathfrak{P}$, $\xi^*\to\gamma_-(1/G_\Gamma)\xi^*$ maps $\mathfrak{K}_0^*$ onto $\mathfrak{K}_\Gamma^*$ and $\mathfrak{K}_0^{*(S)}$ into $\mathfrak{K}_\Gamma^{*(S)}$. The matrix, C_Γ, of the corresponding mapping of $\mathfrak{K}_0^*/\mathfrak{K}_0^{*(S)}$ onto $\mathfrak{K}_\Gamma^*/\mathfrak{K}_\Gamma^{*(S)}$ relative to the obvious bases is given explicitly by

$$(C_\Gamma)_{u,v}=\left\langle \gamma_-\frac{1}{G_\Gamma}\xi^*_{u,0},\ \pi^{v_0}X^v\right\rangle, \tag{9}$$

for all $u, v\in\mathcal{A}'$, $\xi^*_{u,0}$ being as in § 3. Clearly $(C_\Gamma)_{u,v}$ is a power series in Γ which has coefficients which are independent of p and which lie in the field obtained by adjoining the coefficients of f_Γ to the field of rational numbers.

By differentiating both sides of equation (9) we readily establish that

$$\frac{dC_\Gamma}{d\Gamma}=C_\Gamma B \quad (C_0=I), \tag{10}$$

where B is an $N\times N$ matrix whose coefficients are rational functions of Γ whose poles lie among the zeros of $g(\Gamma)R(\Gamma)$. (Explicitly $B_{u,v}=\langle\xi^*_{u,\Gamma}, \pi^{v_0+1}X_0X^v h(X)\rangle$.)

We now note that for $\operatorname{ord}\Gamma>0$ we have

$$(\beta^*_\Gamma)=C_{\Gamma^q}^{-1}(\beta^*_0)C_\Gamma. \tag{3}$$

For $\Gamma\in W$, let $\mathfrak{W}_\Gamma=L(b)/\sum_{i=1}^{n+1}D_{i,\Gamma}L(b)$ with $b=(p-1)/p$, and let $\mathfrak{W}_\Gamma^S$ be the subspace of $\mathfrak{W}_\Gamma$ defined as in § 1. Let α_Γ be the endomorphism $\xi\to\psi(\xi F(X,\Gamma))$ of $L(b)$, let $\bar\alpha_\Gamma$ be the mapping of $\mathfrak{W}_\Gamma$ onto $\mathfrak{W}_{\Gamma^q}$ obtained from α_Γ by passage to quotients and let β_Γ be the mapping of $\mathfrak{W}_\Gamma^S$ onto $\mathfrak{W}_{\Gamma^q}^S$ obtained from $\bar\alpha_\Gamma$ by restriction. The basis of $\mathfrak{W}_\Gamma^S$ is taken to be the images in that space of the element $\{\pi^{u_0}X^u\}_{u\in\mathcal{A}'}$.

For $\operatorname{ord}\Gamma>0$ we now have the diagram

$$\begin{tikzcd}
\mathfrak{K}^*_{\Gamma^q}/\mathfrak{K}^{*(S)}_\Gamma \arrow[r, "C^{-1}_{\Gamma^q}"] \arrow[d, "(\beta^*_\Gamma)"'] & \mathfrak{K}^*_0/\mathfrak{K}^{*(S)}_0 \arrow[r, "J"] \arrow[d, "(\beta^*_0)"'] & \mathfrak{W}^S_0 \arrow[r, "(C^t_{\Gamma^q})^{-1}"] \arrow[d, "(\beta_0^{-1})"'] & \mathfrak{W}^S_{\Gamma^q} \arrow[d, "(\beta_\Gamma^{-1})"'] \\
\mathfrak{K}^*_\Gamma/\mathfrak{K}^{*(S)}_\Gamma \arrow[r, "C^{-1}_\Gamma"'] & \mathfrak{K}^*_0/\mathfrak{K}^{*(S)}_0 \arrow[r, "J"'] & \mathfrak{W}^S_0 \arrow[r, "(C^t_\Gamma)^{-1}"'] & \mathfrak{W}^S_\Gamma
\end{tikzcd}$$

where the mappings are denoted by matrices. The diagram is commutative except in the middle square (§ 3) and we conclude that

$$q^{n+1}(\beta_\Gamma)^{-1}=M_{\Gamma^q}(\beta^*_\Gamma)M_\Gamma^{-1}, \tag{11}$$

where $M_\Gamma = C_\Gamma^t J C_\Gamma$. Since (β_Γ^*) and $(\beta_\Gamma)^{-1}$ are meromorphic on W, Krasner's uniqueness theorem shows that (11) will be valid for all but a finite set of points of W, if M_Γ is meromorphic on W. This finite set of points would cause no difficulty and thus we may conclude that if M_Γ is meromorphic in W then the functional equation is valid $(p \neq 2,\ p \nmid d)$ for all hypersurfaces defined by $f_\Gamma(X) \equiv 0 \bmod \mathfrak{P}$ for some $\Gamma \in W$.

We note that if M is rational then the classical theory of ordinary differential equations shows that M and M^{-1} have poles only at the zeros of $g(\Gamma)R(\Gamma)$.

§ 5. Elliptic curves

It is well known that a non-singular cubic curve defined over the field of q elements has zeta function $(1-\omega_1 t)(1-\omega_2 t)/((1-t)(1-qt))$, that $\omega_1\omega_2 = q$, that if the Hasse invariant of the curve is zero then $\omega_i = \pm\sqrt{-q}$ and if the Hasse invariant is not zero then of the two roots, ω_1, ω_2, one is a p-adic unit and the other is (clearly) a non-unit.

We now refer to an unpublished result of Tate and Lazard.(1)

"Let k be a field of characteristic zero, complete with respect to a discrete valuation, with valuation ring $\mathfrak{o}$ and residue class field $\bar{k} = \mathfrak{o}/\mathfrak{p}$. Let A be an elliptic curve defined over k by an equation

$$Y^2 + (a_1 X + a_2) Y = X^3 + a_3 X^2 + a_4 X + a_5,$$

where the $a_i \in \mathfrak{o}$ and let $\mathrm{du} = -dX/(2Y + a_1 X + a_2)$ be the differential of the first kind on A. In terms of the uniformizing parameter $t = X/Y$ at infinity, we have after integration

$$u = t + \frac{1}{2} D_1 t + \frac{1}{3} D_2 t^3 + \ldots,$$

where the D_i lie in $\mathfrak{o}$.

THEOREM. 1. *If the reduced curve $\bar{A}$ defined over $\bar{k}$ is non-singular and has Hasse invariant not zero then there exists a unit C in the maximal unramified extension, K, of k such that* $\exp(Cu)$ *(as power series in t) has integral coefficients in K.*

2. *If $\bar{k}$ is finite, i.e. k is a p-adic number field then C satisfies the condition*

$$C^{\sigma-1} = \omega,$$

where ω is the unit root of the zeta function of the reduced curve $\bar{A}$ and σ is the Frobenius automorphism of K over k".

If for $p \neq 2$ the equation of A is in the Legendre normal form

$$Y^2 = X(X-1)(X-\Gamma)$$

then the Hasse invariant is $G(\Gamma) \bmod p$, where

$$G(\Gamma) = \sum_{j=0}^{(p-1)/2} \binom{-\frac{1}{2}}{j}^2 \Gamma^j.$$

(1) We quote (with some paraphrasing) a letter of J. Tate to the author (2/13/58). We understand that the statements were conjectured by Tate and proven by him and Lazard using the theory of formal Lie algebras. We are not familiar with the proofs.

For each integer $m \geqslant 0$, let

$$g(2m, \Gamma) = \sum_{i=0}^{m} \binom{-\frac{1}{2}}{i} \binom{-\frac{1}{2}}{m-i} \Gamma^j$$

and let

$$U(\Gamma) = \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j}^2 \Gamma^j, \quad \text{the hypergeometric series } F(\tfrac{1}{2}, \tfrac{1}{2}, 1, \Gamma).$$

The first assertion of Tate and Lazard may be deduced for this case with the aid of [4, Lemma 1] by use of the non-trivial congruence,

$$g(mp^{s+1}-1, \Gamma)\, U(\Gamma^p) - U(\Gamma)\, g(mp^s - 1, \Gamma^p) \in p^{s+1}\{\Gamma\} \tag{12}$$

for all $s \in Z^+$, m odd and positive.

If we let $W = \{\Gamma \in \mathfrak{D} \mid \operatorname{ord} G(\Gamma) = 0\}$ then we may conclude from (12) that the function $H(\Gamma)$ defined by the power series $U(\Gamma)/U(\Gamma^q)$ for $\operatorname{ord} \Gamma > 0$ has an analytic continuation to a holomorphic function on W (again denoted by H).

The second result of Tate and Lazard shows that if $\Gamma \in W$, $\Gamma^q = \Gamma$, $\Gamma \neq 0, 1$ then

$$(-1)^{(q-1)/2} H(\Gamma)$$

is the unit root of the zeta function of the corresponding reduced elliptic curve.

There may be an explanation of the factor $(-1)^{(q-1)/2}$. The zeta function of the singular cubic obtained by setting $\Gamma = 0$ and reducing mod p, is easily found to be

$$\frac{(1-(-1)^{(q-1)/2} t)}{(1-t)(1-qt)}$$

and hence we may say that $(-1)^{(q-1)/2}$ is the unit root of this zeta function. The results for the family of Legendre normal form are now formally similar to equation (3). Aside from the suggestion of a theory for non-regular curves, the results of this section may suggest the correct interpretation of the matrix C in § 4.

§ 6. Newton polygon of $P(t)$

In this paragraph we shall use the notation of [3, § 4].

Let $\mathfrak{B}$ be a subset of $\mathfrak{T}$ such that $\{X^u\}_{u \in \mathfrak{B}}$ is a basis of V. We assume the construction of $F(X)$ (and hence of α) in equation (4.21) and (2.23) of [3] is done with $s = \infty$. Using Lemma 3.6 [3] it is easy to show that the matrix of $\bar{\alpha}$ relative to this basis has the property

$$\operatorname{ord} (\bar{\alpha})_{u,v} \geqslant \frac{p v_0 - u_0}{p-1}$$

(provided H is regular and defined over the prime field). The same statement holds for $\bar{\alpha}_s^s$ if $p \nmid d$. The conclusions concerning Newton polygons

deduced from this statement are valid without the hypotheses in parenthesis. The removal of these restrictive hypotheses is done by the methods of [3, § 4, c] and the methods used in the proof of Theorem 4.3 [3].

Let γ_j be the coefficient of t^j in the polynomial $[t(1-t^{d-1})/(1-t)]^{n+1}$. Let γ_j' be the coefficient of t^j in the polynomial $([1-t^d]/[1-t])^{n+1}$. For $j=0, 1, \ldots, n$ let Q_j be the point $(\sum_{i=0}^{j} \gamma_{id}, \sum_{i=0}^{j} (i-1)\gamma_{id})$ in two dimensional real space and let Q_j' be given by the same formula with γ_{id} replaced by γ_{id}'. Let $\mathfrak{X}$ be the finite convex polygon with vertices $Q_0, Q_1, \ldots, Q_n$; and let $\mathfrak{X}'$ be the finite convex polygon with vertices $Q_0', Q_1', \ldots, Q_n'$ and the origin.

In the following the construction of Newton polygons will be normalized by taking the valuation with ord $q=1$.

THEOREM. (i) *If* $1=(2, p, d)$ *then the non-vertical sides of the Newton polygon of* $P=P_S$ *lie in* $\mathfrak{X}$.

(ii) *If* $\mathfrak{H}$ *is regular then the non-vertical sides of the Newton polygon of* $(1-q^{-1}t)\Pi_A P_A(t)$ *lie in* $\mathfrak{X}'$.

(iii) *If* $1=(2,p,d)$ *then the* p*-adic valuation of the coefficient of* t^N *in* $P(t)$ *is the same as that predicted by conjecture* II.

To aid in possible generalizations of this theorem, it is of interest to identify the numbers γ_{jd}. With the help of the explicit formulae of Hirzebruch [6] it can be verified that

$$\gamma_{(j+1)d}=\begin{cases} h^{j,(n-1)-j} & \text{if } 2j\neq n-1, \\ h^{j,j}-1 & \text{if } 2j=n-1, \end{cases}$$

where $h^{p,q}$ is the "Hodge number" associated with a non-singular hypersurface of degree d and complex dimension $n-1$ in complex projective space. Finally we note that in the notation of [3], γ_{jd} is the dimension of the image in $\mathfrak{W}_S^S$ of polynomials $X_0^j g$, where g runs through all forms of degree dj in the principal ideal $(\Pi_{i=1}^{n+1} X_i)$ in the ring of polynomials in $n+1$ variables.

REFERENCES

[1]. WEIL, A., Numbers of solutions of equations in finite fields. *Bull. Amer. Math. Soc.*, 55 (1949), 497–508.

[2]. LANG, S., *Abelian Varieties*. Interscience Tracts, No. 7, New York, 1959.

[3]. DWORK, B., On the zeta function of a hypersurface. *Pub. Math. (I.H.E.S.), Paris*, 12 (1962).

[4]. —— On the rationality of the zeta function of an algebraic variety. *Amer. J. Math.*, 82 (1960), 631–648.

[5]. SERRE, J. P., Endomorphisms complètement continus des espaces de Banach p-adiques. *Pub. Math. (I.H.E.S.), Paris*, 12 (1962).

[6]. HIRZEBRUCH, E., Neue topologische Methoden in der algebraischen Geometrie. *Proc. Internat. Congress, Amsterdam*, 1954, III, 457–473.

[7]. KRASNER, M., Théorie Des Corps Valués. *C. R. Acad. Sci., Paris*, 238 (1954), 2385–2387; 239 (1954), 468–470, 745–747; 244 (1957), 1304–1306, 1599–1602, 1996–1999, 2570–2573; 245 (1957), 1285–1288.

EINFACH ZUSAMMENHÄNGENDE ALGEBRAISCHE GRUPPEN IN DER ARITHMETIK

Von MARTIN KNESER

Viele Probleme der algebraischen Zahlentheorie lassen sich als Fragen über algebraische Gruppen formulieren. Für gewisse Fragen spielen dabei die einfach zusammenhängenden Gruppen eine besondere Rolle, d. h. diejenigen (im Sinne der Zariski-Topologie) zusammenhängenden algebraischen Gruppen G, für die es keinen surjektiven Homomorphismus $H \to G$ einer zusammenhängenden algebraischen Gruppe H mit endlichem von 1 verschiedenem Kern gibt. Das soll an zwei Beispielen, den Fragen der Approximation und der homogenen Räume dargelegt werden. Zur Vereinfachung der Bezeichnung beschränken wir uns dabei auf die rationalen Zahlen Q als Grundkörper; über algebraischen Zahlkörpern gelten ganz entsprechende Ergebnisse. Außerdem setzen wir voraus, daß G halbeinfach ist; der allgemeine Fall läßt sich ohne weiteres auf diesen zurückführen.

Um bei den verschiedenen Resultaten angeben zu können, in welchen Fällen sie gesichert sind, muß zunächst etwas über die Klassifikation halbeinfacher Gruppen gesagt werden. Über der algebraisch abgeschlossenen Hülle $\overline{Q}$ von Q ist jede solche Gruppe bis auf Isogenie (d. h. bis auf Homomorphismen mit endlichem Kern) isomorph einem Produkt von einfachen Gruppen, und die einfachen Faktoren werden durch die Einteilung in die bekannten Klassen $A_n, \ldots, G_2$ klassifiziert. Über Q selbst zerfällt jede halbeinfache Gruppe bis auf Isogenie in über Q definierte Gruppen, die sich nicht mehr weiter in dieser Weise zerlegen lassen. Diese „Q-einfachen" Faktoren zerfallen über $\overline{Q}$ in lauter isomorphe Faktoren. Falls diese zu einem der Typen $A_n, \ldots, D_n$ gehören, so hat man (abgesehen von D_4) eine Klassifizierung bis auf Isogenie über Q [15]. Es ergeben sich gerade die „klassischen Gruppen", und zwar für den Typ

A_n: die multiplikative Gruppe der Elemente der reduzierten Norm 1 einer einfachen Algebra über Q und die spezielle unitäre Gruppe einer hermiteschen Form über einem Schiefkörper über Q bezüglich eines involutorischen Antiautomorphismus, der Q aber nicht das ganze Zentrum elementweise fest läßt;

B_n: die spezielle orthogonale Gruppe einer quadratischen Form in einer ungeraden Anzahl von Variablen über einem endlichen Erweiterungskörper von Q;

C_n: die symplektische Gruppe und die unitäre Gruppe einer hermiteschen Form über einer Quaternionenalgebra über einer endlichen Erweiterung von Q bezüglich der Standardinvolution;

D_n: die spezielle orthogonale Gruppe einer quadratischen Form in einer geraden Anzahl von Variablen und die spezielle unitäre Gruppe einer schief-hermiteschen Form über einer Quaternionenalgebra bezüglich der Standardinvolution.

Zu diesen Gruppen kommen noch die vom Phänomen der Trialität herrüh-

renden zum Typ D_4 gehörigen Gruppen und die Gruppen der Ausnahmetypen E_6, E_7, E_8, F_4, G_2.

Bei einfach zusammenhängenden Gruppen kann man den Zusatz „bis auf Isogenie" weglassen, nur hat man dann die unter B_n und D_n genannten Gruppen durch ihre einfach zusammenhängenden zweiblättrigen Überlagerungen zu ersetzen, die orthogonalen Gruppen also z. B. durch die zugehörigen Spingruppen.

Approximationssätze. G sei eine über Q definierte algebraische Matrizengruppe; ist K irgendein Integritätsbereich der Charakteristik 0 (z. B. $K=Z_p$ ganze p-adische Zahlen, $K=Q_p$ rational p-adische Zahlen, $K=R$ reelle Zahlen), so sei $G(K)$ die Gruppe der Matrizen X aus G, deren Komponenten ebenso wie die der Inversen X^{-1} in K liegen. Weiter sei S eine endliche Menge von Primzahlen, für jedes p aus S sei X_p ein Element aus $G(Q_p)$ und e_p eine ganze Zahl. Wir sagen, für G gelte der starke Approximationssatz, wenn es für beliebige S, X_p, e_p stets eine Matrix X in $G(Q)$ gibt mit

$$\begin{array}{ll} X \equiv X_p \bmod p^{e_p} & \text{für } p \in S, \\ X \in G(Z_p) & \text{für } p \notin S, \end{array}$$

Satz 1. *Wenn für G der starke Approximationssatz gilt, dann ist G einfach zusammenhängend und $G(R)$ nicht kompakt.*

Die zweite Bedingung ergibt sich durch einfache topologische Schlüsse aus der Deutung des Approximationssatzes als Dichtigkeitsaussage in der Adelgruppe von G; der Beweis der ersten Bedingung ist schwieriger und benützt u. a. das Resultat aus [1], daß $G(Z)$ endlich erzeugt ist.

Fragt man umgekehrt nach hinreichenden Bedingungen, so kann man sich auf halbeinfache Gruppen beschränken, da der Approximationssatz für eine einfach zusammenhängende Gruppe G genau dann gilt, wenn er für den halbeinfachen Teil von G gilt; außerdem braucht man nur Q-einfache Gruppen zu betrachten, da der Approximationssatz für ein direktes Produkt genau dann gilt, wenn er für alle Faktoren gilt. Hierfür hat man

Satz 2. *Ist G eine einfach zusammenhängende Q-einfache klassische Gruppe (ausgenommen vielleicht die unitären Gruppen vom Typ A_n, $n \geqslant 2$) und $G(R)$ nicht kompakt, oder eine einfach zusammenhängende Gruppe von dem von Chevalley* [3] *behandelten Typ, so gilt der starke Approximationssatz.*

Dieses Ergebnis stammt für den Typ A_n von Eichler [5]; für die anderen klassischen Gruppen benützt man in niedrigen Dimensionen die bekannten Isomorphismen mit Gruppen vom Typ A_1, in höheren Dimensionen die Existenz hinreichend vieler Untergruppen niedrigerer Dimension, für die der Approximationssatz also schon bewiesen ist. Bei den Chevalleyschen Gruppen schließlich zieht man die Existenz hinreichend vieler über Q rationaler unipotenter Elemente heran.

Als Anwendungen erhält man die Bestimmung verschiedener Klassenzahlen, z. B. Idealklassen von Hauptordnungen indefiniter Schiefkörper [4], oder Klassen indefiniter quadratischer Formen [9] (der dortige kompliziertere Approximationssatz für orthogonale Gruppen ist eine einfache Folge des starken Approximationssatzes für die zugehörigen Spingruppen), sowie Ergebnisse im Zusammenhang mit Siegels Hauptsatz über quadratische Formen [6, 10].

Homogene Räume. Sei G eine algebraische Gruppe, H ein homogener G-Raum, d. h. eine algebraische Mannigfaltigkeit, auf der G als algebraische Transformationsgruppe transitiv wirkt. H heißt prinzipaler homogener Raum, wenn für x und y aus H dasjenige g aus G, welches $y=g(x)$ erfüllt, eindeutig bestimmt und rational in x und y ist. Verschiedene Probleme, z. B. die Klassifikation von einfachen Algebren, von quadratischen Formen usw., führen auf die Aufgabe, die über einem Körper K definierten prinzipalen homogenen Räume einer algebraischen Gruppe bezüglich K-Isomorphie zu klassifizieren. Es ist bekannt, daß über einem endlichen Körper K jeder homogene Raum einer zusammenhängenden algebraischen Gruppe einen über K rationalen Punkt hat, daß also insbesondere alle prinzipalen homogenen Räume K-isomorph sind [12]. Über $\mathfrak{p}$-adischen Körpern, d. h. vollständigen bewerteten Körpern der Charakteristik 0 mit endlichem Restklassenkörper, und über Zahlkörpern sind die Verhältnisse komplizierter.

Satz 3. *Ist K ein $\mathfrak{p}$-adischer Körper, G eine einfach zusammenhängende klassische Gruppe, H ein prinzipaler homogener G-Raum, beide über K definiert, so hat H einen über K rationalen Punkt.*

Über Q (und entsprechend über Zahlkörpern) sind eine Reihe von Fällen bekannt, in denen ein prinzipaler homogener G-Raum genau dann einen über Q rationalen Punkt hat, wenn dies über allen Q_p und über R gilt. Für die orthogonalen Gruppen ist dies z. B. gerade der Satz von Hasse über quadratische Formen [7], für die projektive lineare Gruppe der von R. Brauer, H. Hasse und E. Noether über den Zerfall einfacher Algebren [2]. Ein Beispiel von Serre zeigt aber, daß dies nicht für alle zusammenhängenden halbeinfachen Gruppen gilt. Für einfach zusammenhängende Gruppen fallen nach Satz 3 die Bedingungen für die Q_p weg. Jeden über Q definierten homogenen Raum kann man nun auch als über R definiert auffassen und erhält so eine Abbildung von Q-Isomorphieklassen homogener Räume auf R-Isomorphieklassen. Hier gilt nun nicht nur eine Aussage über die Existenz rationaler Punkte, sondern etwas genauer sogar

Satz 4. *Ist G eine einfach zusammenhängende über Q definierte klassische Gruppe, so ist die angegebene Zuordnung von Q-Isomorphieklassen prinzipaler homogener G-Räume zu R-Isomorphieklassen bijektiv.*

Für den ersten unter A_n genannten Typ ist dieser Satz der Normensatz für einfache Algebren [8, 13]; für den zweiten Typ A_n ist er im wesentlichen äquivalent zu Ergebnissen von Landherr [11]. Für die übrigen Gruppen benutzt man zum Beweis wieder in niedrigen Dimensionen die bekannten Isomorphismen und dann Induktion nach der Dimension.

Über prinzipale homogene Räume zusammenhängender aber nicht einfach zusammenhängender Gruppen G erhält man Auskunft, indem man die Isomorphieklassen prinzipaler homogener G-Räume als eindimensionale Galois-Kohomologieklassen deutet und dann die zu einer exakten Folge $1\to E\to F\to G\to 1$ mit endlichem E und einfach zusammenhängendem F gehörige exakte Kohomologiefolge ausnützt. Entsprechend kann man Fragen über nicht zusammenhängende Gruppen G mittels der exakten Folge $1\to G_0\to G\to G/G_0\to 1$ auf solche über die Zusammenhangskomponente G_0 der Eins und über G/G_0 zurückführen. Für die Behandlung prinzipaler

homogener Räume mit Galois-Kohomologie und damit zusammenhängende Fragen vergleiche man den Bericht [14].

Zum Schluß sei darauf hingewiesen, daß es sehr erwünscht wäre, für die hier angegebenen Sätze Beweise zu finden, welche alle Gruppen gemeinsam behandeln und keine Fallunterscheidungen nötig machen.

LITERATUR

[1]. BOREL, A. & HARISH-CHANDRA, Arithmetic subgroups of algebraic groups. *Ann. Math.*, 75 (1962), 485–535.

[2]. BRAUER, R., HASSE, H. & NOETHER, E., Beweis eines Hauptsatzes in der Theorie der Algebren. *J. reine angew. Math.*, 167 (1932), 399–404.

[3]. CHEVALLEY, C., Sur certains groupes simples. *Tôhoku Math. J.*, 7 (1955), 14–66.

[4]. EICHLER, M., Über die Idealklassenzahl hyperkomplexer Systeme. *Math. Z.*, 43 (1938), 481–494.

[5]. —— Allgemeine Kongruenzklasseneinteilungen der Ideale einfacher Algebren über algebraischen Zahlkörpern und ihre L-Reihen. *J. reine angew. Math.*, 179 (1938), 227–251.

[6]. —— Die Ähnlichkeitsklassen indefiniter Gitter. *Math. Z.*, 55 (1952), 216–252.

[7]. HASSE, H., Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen. *J. reine angew. Math.*, 152 (1923), 205–224.

[8]. HASSE, H. & SCHILLING, O., Die Normen aus einer normalen Divisionsalgebra über einem algebraischen Zahlkörper. *J. reine angew. Math.*, 174 (1936), 248–252.

[9]. KNESER, M., Klassenzahlen indefiniter quadratischer Formen in drei oder mehr Veränderlichen. *Arch. Math.*, 7 (1956), 323–332.

[10]. —— Darstellungsmaße indefiniter quadratischer Formen. *Math. Z.*, 77 (1961), 188–194.

[11]. LANDHERR, W., Liesche Ringe vom Typus A über einem algebraischen Zahlkörper und hermitesche Formen über einem Schiefkörper. *Abh. Math. Sem. Hamburg*, 12 (1938), 200–241.

[12]. LANG, S., Algebraic groups over finite fields. *Amer. J. Math.*, 78 (1956), 555–563.

[13]. MAASS, H., Beweis des Normensatzes in einfachen hyperkomplexen Systemen. *Abh. Math. Sem. Hamburg*, 12 (1938), 66–69.

[14]. SERRE, J.-P., Cohomologie galoisienne des groupes algébriques linéaires, *Coll. Théorie des Groupes Algébriques*. Bruxelles, 1962.

[15]. WEIL, A., Algebras with involutions and the classical groups, *J. Ind. Math. Soc.*, 24 (1961), 598–623.

АЛГЕБРЫ ЛИ И КОНЕЧНЫЕ ГРУППЫ

А. И. КОСТРИКИН

1. Метод алгебр Ли в теории конечных групп и структурные свойства простых алгебр Ли над алгебраически замкнутым полем конечной характеристики — таковы две темы, которых мне надлежит коснуться в настоящем докладе. Связь между алгебрами Ли и конечными группами уже была предметом обсуждения на международном коллоквиуме по конечным группам в Тюбингине в 1957 году, на международном математическом конгрессе в Эдинбурге (Шевалле [1], Хигман [3]) и на симпозиуме по конечным группам в Нью-Йорке в 1959 году (Линдон [11], Цасенхауз [18]). Один из мотивов для выбора столь популярной сейчас темы заключается в том, что известно несколько глубоких результатов о конечных группах, получение которых без аппарата алгебр Ли пока не представляется возможным. В докладе затрагивается лишь небольшая часть их. Круг рассматриваемых вопросов ограничен, в основном, теми результатами и задачами, которые объединяются работами автора по проблеме Бернсайда.

2. Я начну с замечания о том, что в основе многих плодотворных исследований, связанных с простыми конечными группами (работы Шевалле, Титса, Стейнберга, Херцига, Ри и др.), лежит тщательно разработанная ранее теория простых алгебр Ли над полем комплексных чисел. Джекобсон перенес основные результаты структурной теории на алгебры Ли, определенные над произвольным алгебраически замкнутым полем характеристики нуль. Было бы чрезвычайно интересно, как с точки зрения возможных приложений, так и из естественного стремления иметь законченную алгебраическую картину, получить классификационные теоремы для простых алгебр Ли конечной характеристики.

Исследование алгебр Ли (как, впрочем, и многих других алгебраических объектов) над полями конечной характеристики сопряжено с большими трудностями. Наши сведения о них, несмотря на значительный вклад Цасенхауза [16, 17], Джекобсона [4–6], Капланского [7] и других авторов, далеки от того, чтобы быть полными. Одна из причин такого положения вещей заключается в том, что стержень классической теории — признак полупростоты Картана — теряет свою силу в случае характеристики $p>0$. А именно, существуют простые алгебры Ли с фундаментальными билинейными формами $B(x,y)=TrXY$, тождественно равными нулю (большими буквами $X, Y, \ldots$ обозначаются присоединенные эндоморфизмы $z\to[zx], z\to[zy], \ldots$ векторного пространства $\mathfrak{L}$, соответствующие элементам $x, y, \ldots$ из алгебры $\mathfrak{L}$).

В дальнейшем такие алгебры будут называться вырожденными. Это весьма обширный класс алгебр, полный загадок и нерешенных вопросов. Наиболее ранним примером простой вырожденной алгебры Ли является алгебра Витта $\mathfrak{L}=\{e_0, e_1, \ldots, e_{p-1}\}$, $[e_i e_j]=(j-i)\, e_{i+j}, p>3$. В настоящее время известно несколько бесконечных серий вырожденных простых алгебр Ли размерностей mp^n, $1\leqslant m\leqslant n$, $m(p^n-1)$, $1\leqslant m\leqslant n-1$,

p^n-2 и др. Некоторые серии состоят из p-алгебр Ли (или алгебр, ограниченных в смысле Джекобсона), являющихся по многим причинам наиболее естественным объектом исследования в случае конечной характеристики. Для определенности ниже речь будет идти о p-алгебрах, хотя это ограничение и не всегда существенно. Используя, в основном, классические методы, Селигман [14] описал все простые p-алгебры Ли с невырожденной билинейной формой следа $Tr\varrho(x)\varrho(y)$, соответствующей какому-нибудь линейному представлению ϱ. Они аналогичны простым алгебрам Ли над полем комплексных чисел и характеризуются известными свойствами своих картановских разложений. Эти свойства почти полностью утрачиваются в вырожденных простых p-алгебрах Ли, теряя свой инвариантный характер. Здесь будет доложено об одном новом подходе к исследованию вырожденных простых p-алгебр Ли, который представляется не лишенным смысла хотя бы потому, что он, не претендуя на универсальность, заменяет весьма неприятный факт исчезновения метрики (фундаментальной билинейной формы $B(x,y)$) на одно очень прозрачное алгебраическое условие. По причинам, которые скоро будут ясны, удобно, однако, прервать изложение о простых алгебрах, рассказав предварительно о совершенно другой задаче.

3. Имеется в виду ослабленная проблема Бернсайда для простого показателя p.

Существует ли в классе всех конечных групп с заданным числом k образующих и тождественным соотношением $x^p=1$ такая максимальная группа $\bar{B}_{k,p}$, что любая другая группа из этого класса является ее фактор-группой?

В настоящее время известен положительный ответ на этот вопрос [8].

Он получается в качестве следствия следующей теоремы. Произвольное кольцо Ли L характеристики p, удовлетворяющее n-му условию Энгеля, $n\leqslant p$ (n-любое, если $p=0$), — локально нильпотентно.

Условие Энгеля выражается в требовании, чтобы $[uv^n]=[\ldots[[uv]v]\ldots v]=0$ для всех элементов u,v из L. Здесь нет необходимости (как, впрочем, и возможности) пояснять, каким образом устанавливается соответствие между группами и кольцами Ли; это довольно хорошо известно и излагалось, например, в докладе Хигмана на прошлом конгрессе. Мне хотелось бы лишь сделать несколько замечаний о группах с тождественным соотношением $x^p=1$.

Пусть F — свободная группа с k образующими, F_n — члены ее нижнего центрального ряда: $F_1=F$, $F_2=(F,F)$, ... $F_n=(F_{n-1},F)$. Обозначим через F^p характеристическую подгруппу, порожденную p-ми степенями всех элементов из F, и положим $B_{k,p}=F/F^p$.

Входящая в формулировку ослабленной гипотезы Бернсайда группа $\bar{B}_{k,p}$ существует в том и только в том случае, если найдется такое (наименьшее) целое положительное число $c=c(k,p)$, что $F_{c+i}\,F^p=F_c\,F^p$ для всех i. Коль скоро это так, $\bar{B}_{k,p}=F/F_c\,F^p$.

В 1959 году П. С. Новиков [12] решил в отрицательном смысле проблему Бернсайда о периодических группах, доказав существование бесконечной конечно-порожденной группы с тождественным соотношением $x^n=1$, где n — достаточно большое целое число.

В случае $n=p$ это означает попросту, что включения $F_i\subset F^p$ нет ни для какого индекса i. Легко видеть при этом, что некоторая фактор-

группа группы $F_c F^p/F^p$ (а, возможно, она сама) является простой бесконечной p-группой без подгрупп конечного индекса. Можно добавить еще, что группа $B_{k,p}$ не является энгелевой [9], т. е. не существует числа m такого, что $(g,h;\ m)=1$ для всех $g,h \in B_{k,p}$. Здесь $(g,h)=ghg^{-1}h^{-1}$, $(g,h;\ m)=((g,h;\ m-1),h)$. С другой стороны, группа $\bar{B}_{k,p}$ — энгелева, как всякая конечная p-группа. Показатель энгелевости n_p — минимальное натуральное число, при котором $(g,h;\ n_p)=1$ для всех $g,h \in \bar{B}_{k,p}$ — однозначно определяется простым числом p. Получена оценка $n_p \geqslant (3p-5)/2$ для всех $p>3$. Отсюда, во всяком случае, следует, что хотя ассоциированное с группой $\bar{B}_{k,p}$ кольцо Ли удовлетворяет $(p-1)$-му условию Энгеля, для самой группы это неверно. Доказано равенство $n_5=6$ [9]. Поэтому приобретает интерес следующий вопрос, имеющий непосредственное отношение к проблеме Бернсайда для показателя 5. Выполняется ли тождество $(a,b;\ 6)\equiv 1\,(F^5)$ в свободной группе F с двумя свободными образующими a,b?

4. Я перехожу теперь к очень краткому изложению доказательства теоремы о локальной нильпотентности энгелевых колец Ли с тем, чтобы получить некоторые детали, необходимые в дальнейшем. Постановка вопроса, подлежащего решению, характерна для задач бернсайдовского типа, получивших довольно широкое распространение во всей современной алгебре. Для доказательства в кольце L рассматривается радикал $N(L)$ — сумма локально нильпотентных идеалов. Если предполагать кольцо L, удовлетворяющее n-му условию Энгеля, не локально нильпотентным, то таким же будет и фактор кольцо $L/N(L)$. Можно поэтому с самого начала понимать под L ненулевое кольцо без радикала $L\neq 0$, $N(L)=0$. В частности, L не имеет центра. Будем говорить далее, что элемент $c\neq 0$ из кольца L принадлежит множеству $L(m)$, если имеют место тождества

$$CX^iC=0 \quad (i=0,1,\ldots,2m-1),$$

для всех элементов $x\in L$.

Очевидно $L(1)\supset L(2)\supset \ldots \supset L(m)\supset \ldots$. Сначала устанавливается непустота множества $L(1)$, т. е. существование элемента $c\neq 0$ индекса нильпотентности два ($C^2=0$), а затем доказывается, что и множество $L((p-3)/2)$ также непусто, что, как нетрудно заметить, ведет к противоречию с нашим предположением относительно кольца L. Весь этот процесс совершается последовательным конструированием элементов множества $L(m+1)$ из элементов множества $L(m)$. Операции при $m\geqslant 2$ являются совершенно формальными и осуществляются в любом кольце Ли без центра. Условие Энгеля используется в полной мере лишь в доказательстве существования элемента $c\neq 0$ ($C^2=0$) и при построении множества $L(2)$. Анализ доказательства показывает, однако, что переход от $L(1)$ к $L(2)$ можно совершить также в любой простой алгебре Ли конечной размерности, если только в ней имеет смысл говорить о множестве $L(1)$.

5. Вырожденные простые p-алгебры Ли — как раз тот класс конечномерных алгебр, в котором только что введенные понятия присутствуют в явном виде [10]. Это замечание содержит в себе элемент неожиданности, благоприятный для описанной выше конструкции, которая в

противном случае казалась бы чересчур искусственной. Действительно, энгелевы кольца того типа, к которым она применима, на самом деле, не существуют. Что же касается ослабленной проблемы Бернсайда для простого показателя p, то когда-нибудь она будет решена более удовлетворительным способом, дающим попутно и (не найденную пока) оценку сверху для класса нильпотентности $c(k,p)$.

Для удобства будем говорить, что простая алгебра Ли $\mathfrak{L}$ имеет сильное вырождение, если множество $\mathfrak{L}(1)$ не пусто, т. е. в $\mathfrak{L}$ существует хотя бы один элемент индекса нильпотентности два. Любая простая p-алгебра Ли с сильным вырождением является вырожденной в обычном смысле. Это утверждение доказать совсем нетрудно. Уместно задать теперь следующий вопрос: какую долю в множестве всех вырожденных простых p-алгебр Ли занимают алгебры с сильным вырождением? Нельзя сказать, что они исчерпывают собою все это множество, так как, например, алгебра A_{np-1} всех $np\times np$ матриц с нулевым следом, профакторизованная по одномерному центру, является вырожденной простой p-алгеброй Ли без сильного вырождения. С другой стороны, для значительного количества бесконечных серий вырожденных простых p-алгебр Ли ($p>3$) установлено сильное вырождение. Следует подчеркнуть, что ограничение $p>3$ существенно. Вполне вероятно, что верна следующая гипотеза. Все вырожденные простые p-алгебры Ли, за исключением искаженных вариантов классических, обладают сильным вырождением. Я не могу пока доказать эту гипотезу во всей общности, однако для некоторых классов алгебр, выделяемых, в основном, свойствами своих картановских разложений, она верна. В качестве примера можно отметить тот факт, что простая p-алгебра Ли с картановской подалгеброй, полупростая компонента которой одномерна, обладает сильным вырождением. Некоторые сведения можно получить об алгебрах, у которых тождественно равны нулю k-линейные формы следа, $k\leqslant 4$. Впрочем, трудности, с которыми приходится иметь дело в данной ситуации, могут служить также и некоторым утешением, свидетельствуя о том, что рассматриваемая характеризация вырожденных алгебр не слишком близка к тривиальной. Основная теорема, доказательство которой является модифицированным вариантом доказательства соответствующей теоремы в энгелевых кольцах Ли, заключается в следующем.

Пусть $\mathfrak{L}$ — простая p-алгебра Ли с сильным вырождением. Тогда в $\mathfrak{L}$ существует элемент $c\in\mathfrak{L}((p-3)/2)$.

В качестве непосредственного следствия получается, что в простой p-алгебре Ли с сильным вырождением имеют место тождества $Tr\, X_1X_2\ldots X_k=0$, $1\leqslant k\leqslant p-2$.

Этот результат нельзя улучшить, так как в известных мне вырожденных простых p-алгебрах Ли множество $\mathfrak{L}((p-1)/2)$ пусто.

Нильпотентные элементы индекса два в простых алгебрах Ли играют роль, в каком-то смысле аналогичную роли элементов второго порядка в простых конечных группах, причем свои преимущества имеет и та, и другая сторона. Отметим здесь одно из наиболее приятных алгебраических свойств.

Векторное пространство $\mathfrak{C}$, натянутое на все элементы множества $\mathfrak{L}(1)$, является нильпотентной подалгеброй в $\mathfrak{L}$. Присоединенные эндоморфизмы алгебры $\mathfrak{L}$, соответствующие элементам из $\mathfrak{C}$, все нильпо-

тентны. В центре $\mathfrak{z}(\mathfrak{C})$ подалгебры $\mathfrak{C}$ содержится по меньшей мере один элемент $c \in \mathfrak{L}((p-3)/2)$.

Можно ожидать, что изучение подалгебр типа $\mathfrak{C}$ и способов их вложения в простые алгебры даст много информации о вырожденных алгебрах. Есть основания предполагать, в частности, что нормализатор подалгебры $\mathfrak{C}$ (а, возможно, и ее центра $\mathfrak{z}(\mathfrak{C})$ в алгебре $\mathfrak{L}$ является разрешимой алгеброй «небольшой» коразмерности в $\mathfrak{L}$.

Используя лишь один фиксированный элемент из $\mathfrak{L}((p-3)/2)$, удалось доказать, что вырожденных простых p-алгебр Ли размерности $<p$ не существует, а единственной такой алгеброй размерности p является алгебра Витта. Получены некоторые другие ограничения на размерности вырожденных алгебр.

7. В заключение остановимся на задаче (представляющей независимый интерес), анализ которой приводит естественным образом к вырожденным простым алгебрам. Что можно сказать о строении алгебры Ли конечной размерности, если на ней действует автоморфизм ϕ, не оставляющий на месте ни одного элемента, кроме нуля? Пусть автоморфизм ϕ (назовем его регулярным) удовлетворяет минимальному уравнению степени n. При небольших n без труда проверяется, что соответствующая алгебра Ли является разрешимой. Исследование общего случая немедленно сводится к вопросу о существовании вырожденной простой алгебры Ли с регулярным автоморфизмом. Но вырождение возможно лишь для алгебр, определенных над полем характеристики $p>0$. Таким образом, конечномерная алгебра Ли над полем характеристики нуль, допускающая регулярный автоморфизм, разрешима. Несколько из других соображений тот же результат был получен ранее Патерсоном [13]. Если $p>0$, то можно показать, что простая алгебра Ли с регулярным автоморфизмом степени $n<p$ обладает сильным вырождением. Это обстоятельство позволяет использовать описанную выше технику для доказательства новых результатов. Впрочем, гораздо большее было получено на другом пути. Совсем недавно В. А. Крекнин доказал следующую общую теорему. Алгебра Ли $\mathfrak{L}$ над полем произвольной характеристики, обладающая регулярным автоморфизмом ϕ конечного периода m, разрешима; имеет место неравенство $d<2^{m-1}$, где d — длина производного ряда алгебры $\mathfrak{L}$.

Отметим интересное следствие, вытекающее из этой теоремы. Как известно (Г. Хигман [2], Дж. Томпсон [15]), всякая алгебра Ли (а также группа конечного порядка) с регулярным автоморфизмом простого периода $m=q$ является нильпотентной, причем для класса нильпотентности существует верхняя граница $k(q)$, зависящая только от q. Рост функции $k(q)$ при больших q до последнего времени оставался почти полностью неисследованным. Оказывается, что справедлива оценка $k(q)<(q-1)^d$ (в свою очередь, $d<2^{q-1}$). Хотя эта оценка, несомненно, сильно завышена, она содержит в себе существенные сведения о рассматриваемых классах алгебр Ли и конечных групп. Любопытно, что доказательство последних утверждений совершенно элементарно. В высшей степени интересным представляется получение аналога теоремы, доказанной Крекниным, для конечных групп при произвольном m.

ЛИТЕРАТУРА

[1]. Chevalley, C., *Proc. Internat. Congress Math.* (1958). Cambridge Univ. Press, 1960.
[2]. Higman, G., *J. London Math. Soc.*, 32 (1957), 321–334.
[3]. —— *Proc. Internat. Congress Math. Edinburgh*, 1958, 307–312. Cambridge Univ. Press, 1960.
[4]. Jacobson, N., *Trans. Amer. Math. Soc.*, 50 (1941), 15–25.
[5]. —— *Amer. J. Math.*, 63 (1941), 481–515.
[6]. —— *Amer. J. Math.*, 74 (1952), 357–359.
[7]. Kaplansky, J., *Trans. Amer. Math. Soc.*, 89, 1 (1958), 149–183.
[8]. Кострикин, А. И., *Известия АН СССР*, 23, 1 (1959).
[9]. —— *Доклады АН СССР*, 135, 3 (1960).
[10]. *Труды Матем. института им. В. А. Стеклова*, LXIV (1961), 79–89.
[11]. Lyndon, R. C., *Proc. Symp. Pure Math.*, 1 (1959), 4–14.
[12]. Новиков, П. С., *Доклады АН СССР*, 127, 4 (1959).
[13]. Patterson, E. M., *Quart. J. Math.*, 12, 46 (1961), 127–133.
[14]. Seligman, G. B., *Mem. Amer. Math. Soc.*, 19 (1956).
[15]. Thompson, J., *Proc. Nat. Acad. Sci.*, 45 (1959), 578–581.
[16]. Zassenhaus, H., *Abh. Math. Sem. Hamburg*, 19 (1939), 1–100.
[17]. —— *Proc. Glasgow Math. Assoc.*, 2, 1 (1954), 1–36.
[18]. —— *Proc. Symp. Pure Math.*, 1 (1959), 105–108.

ADDITIVE PROBLEMS AND EIGENVALUES OF THE MODULAR OPERATORS

By JU. V. LINNIK

§ 1. Methods of solution of additive prolems

The methods of solution of the additive problems known up to now can be sorted out essentially into four types:

(I) The quantity of solutions is represented by a Fourier coefficient and expressed by the integral operator $\int_0^1(\ )\ \exp\ (-2\pi i n\alpha)\ d\alpha$ operating on a trigonometric sum (n denotes the represented number).

Many well known problems of "ternary" type were solved by this method.

(II) The method of the Eratosthene's sieve.

This method is applicable to binary problems with "almost prime" numbers. It gives inequalities only; the asymptotic is not given by this method.

(III) The variance method.

This method was worked out in recent years and was applied mostly to the binary problems involving a binary quadratic form. It gives the asymptotic of the solution quantity, for instance, of the Hardy-Littlewood problem: $p+\xi^2+\eta^2=n$, the divisor problems $p-x\,y=1\quad(x\,y\leqslant n)$; $x\,y-x_1\,x_2\ldots x_n=1\ (x\,y\leqslant n)$ (Ju. V. Linnik [14]), $p=\xi^2+\eta^2+1$; $(p\leqslant n)$ (B. M. Bredikin [3]); the generalized Kloosterman problem: $N(\mathfrak{A}_1)+N(\mathfrak{A}_2)=n$ ($N(\mathfrak{A}_i)$ —ideal norms.)

(IV) Reducing the counting of the solution quantity the estimation of eigenvalues for the corresponding semigroups of "discrete operators".

The method is applicable only to the estimation of the coefficients of modulars forms of an integer dimension, in particular, to the theta—functions giving the representation quantity of a number by a positive quadratic form with an even number of variables. The well-known Hecke operators act here as the above-mentioned operators.

This method is remarkable by the property that in the case of the dimension of modular forms equal to -2, in particular, for quaternary quadratic forms, it gives optimal estimates of the remainder terms (M. Eichler [6]). For primitive quaternary quadratic forms $F(x)$ we get the formula (cf. [6]).

$$r_F(n)=E_F(n)+\sigma_F(n), \tag{1.1}$$

where $r_F(n)$ denotes the representation quantity of the number n by the form F, $E_F(n)=c_F\cdot n\sum_{t\mid n}\left(\frac{D}{t}\right)t^{-1}$ (c_F is a constant);

$$|\sigma_F(n)|<\gamma_F(\varepsilon)n^{1/2+\varepsilon}\quad\text{for each } \varepsilon>0. \tag{1.2}$$

The formula is true for all the numbers which are not divisible by a finite quantity of "exclusive" of finding the optimal remainder term may remain unsolved.

(A little more precise estimate: $|\sigma_F(n)|\leqslant\gamma_F\cdot\tau(n)n^{1/2}$; $\tau(n)$, the divisor quantity and some generalizations were indicated by A. N. Andrianov [1], [2].)

§2. The levelling and its connection with the reduction of curves and problems of the type of H. Hasse's hypothesis. A problem on integer matrices

A straightforward application of method (I) begins usually by sorting out the numbers of the segment [0, 1] into two classes.

Putting $\tau = N^b$ where $b<1$ is a convenient constant we develop the numbers $\alpha \in [0, 1]$ into continued fractions, putting:

$$\alpha = \frac{a}{g} + \frac{Z}{g\tau};\ (a, g) = 1;\ g \leqslant \tau;\ |Z| \leqslant 1 \quad \text{for} \quad g \leqslant \psi(N),$$

where $\psi(N)$ is a comparatively slowly increasing function of n, say, n^ε (ε being small) or $(\ln n)^h$ ($h>1$), the segments

$$I_{ag} = \left[\frac{a}{g} - \frac{1}{g\tau}, \frac{a}{g} + \frac{1}{g\tau}\right]$$

do not intersect.

Now, $\sum_{a,g} \int I_{ag}(\ )\exp(-2\pi\, in\, \alpha)\, d\alpha$ forms the "principal term" of the problem while the absolute value of the integral over the complement to $\sum I_{ag}$ is estimated. If it proves to increase more slowly than the principal term as $n \to \infty$, we get the asymptotic of the problem, though the natural problem the norms of the ideals of prescribed classes in two given algebraic number fields) (A. I. Vinogradov [21], etc.).

However, in many additive problems such an approach proves to be not precise enough. Such a situation arises, for instance, in the binary Goldbach's problem: $p_1 + p_2 = n$, or the twin primes problem: $p_1 - p_2 = 2$, $p_1 \leqslant n$ (cf. Ju. V. Linnik [10], [11]). The conditional investigation of these problems taking into account the extended Riemann hypothesis leads to the situation described above.

Thus, we can solve the equation: $n = p_1 + p_2 + 2^{x_1} + \ldots + 2^{x_k}$ (k being a constant); the Riemann hypothesis can be replaced by the proved density theorems (Ju. V. Linnik [11], A. I. Vinogradov [20]).

The general approach mentioned above is not precise enough here.

Such a situation was discovered as far back as 1926 by H. D. Kloosterman [9] by studying the equation (1.1). In the article [9] was given a partition of the segment [0, 1] more sophisticated than the above-described one. The segments I_{ag} which are disjoint for small g's, can intersect for the numbers g and g_1, such that $g + g_1 > \tau$. To get rid of such intersections and at the same time to fill the segment [0, 1] without gaps, the well-known Farey series is applied; instead of the segments I_{ag}, the segments j_{ag} are taken whose length is unequal for different a's:

$$j_{ag} = \left[\frac{a}{g} - \frac{1}{g(g+g_2)}, \frac{a}{g} + \frac{1}{g(g+g_1)}\right],$$

where a_1/g_1 and a_2/g_2 are the neighbours of a/g in the Farey series. These

segments j_{ag} are dissected further into still smaller segments j'_{mg} which are equal length for a given q and equidistant on the segment $:0, 1;$. In the integral $\int_0^1(\ )\exp(-2\pi in\alpha)d\alpha$ giving the solution number, the corresponding quantities are summed up for a given q and the numbers m in the segments j'_{ma} in the natural order; these sums are then estimated by their absolute value. This method, the idea of which can be traced back to G. Hardy, enables to extend the sphere of application of method (I). Its application leads to the appearance in the number theory of H. Kloosterman's sums [9]:

$$S(a, b, g) = \sum_{x \bmod g} \exp\frac{2\pi i}{g}(ax' + bx); \quad (xx' \equiv 1 (\bmod g)) \tag{2.1}$$

(we shall discuss their properties further). A. V. Malyshev developed this method still further [15, 16, 17], using the well-known estimates of A. Weil [18] for trigonometric sums to advance considerably in the problem of optimal estimates of the remainder terms in the theory of quadratic forms. If $F(x_1, \ldots, x_n)$ is an integer quadratic form, then for the solution quantity $r_F(n)$ of the equation

$$F(x_1, \ldots, x_k) = n \tag{2.2}$$

under certain necessary conditions for the numbers n, which are not divisible by certain exclusive prime numbers (in a finite number), A. V. Malyshev gives the expression of the type (2.1):

$$r_F(n) = E_F(n) + \sigma_F(n), \tag{2.3}$$

where $E_F(n)$ is a term analogous to that of formula (1.1) and

$$|\sigma_F(n)| < \gamma_F(\varepsilon) n^{\frac{1}{2}k - \frac{1}{2} + \varepsilon}. \tag{2.4}$$

As is well known, the optimal expected estimate in this problem must be of the type (2.4) with the exponent $\frac{k}{4} - \frac{1}{2} + \varepsilon$, and thus differs from A. V. Malyshev's exponent by the quantity $\frac{1}{4}$. For $k=4$, the formulas of M. Eichler (1.1); (1,2) give the expected estimate for $k>4$ the problem of the optimal extimate remains unsolved. For the particular case $k=24$:

$$F(x_1, \ldots, x_k) = x_1^2 + \ldots + x_{24}^2$$

this question forms essentially the well known Ramanuyan-Eichler problem [7]. The result (2.4) was obtained after carrying out the above-mentioned method of H. Kloosterman, namely to express $\int_0^1(\ )\exp(-2\pi i\alpha)d\alpha$ by means of the integrals extended over the segments [0, 1] of equal length for a given g, equidistant of the segment [0, 1].

Let an additive problem be given with the number of solutions $r(n)$ expressed by the integral:

$$r(n) = \int_0^1 (\ ) \exp(-2\pi in\alpha)d\alpha. \tag{2.5}$$

Suppose we succeeded in replacing with an error (admissible in comparison with the "principal term") $\gamma(n)$ the integral (2.5) by the sum:

$$r_1(n) = \sum_{g \leqslant \sqrt{N}} \sum_{(a,g)=1} I_{ag}, \tag{2.6}$$

where

$$I_{ag} = \int_{j_{ag}} (\ ) \exp(-2\pi i n\alpha) d\alpha \tag{2.7}$$

$$j_{ag} = \left[\frac{a}{g} - \frac{N^\varepsilon}{N}, \frac{a}{g} + \frac{N^\varepsilon}{N}\right];$$

$\varepsilon > 0$ an arbitrarily small constant.

In such a case we shall say that we carried out the levelling in the additive problem (2.5) with the error $\gamma(n)$. The main point of the operation of levelling consists in the condition that for a given g and for all a's under condition $(a,g)=1$, the segments j_{ag} are of equal length; we could admit that this length is not constant for different g's but changes sufficiently slowly if g changes, further, g may be supposed to change till $\sqrt{N}(\ln N)^h$ or N^β (with $\beta < \frac{1}{2}$) instead of $\sqrt{N}$. We remark that the operation of levelling does not exclude the overlapping of the segments j_{ag}.

If the problem (2.5) admits the levelling with the error $\gamma(n)$ in the sense mentioned above or the wider sense mentioned below, the problem then can be reduced to the problem of the asymptotic of solutions of certain congruences.

Denote the term which is to be put into the brackets in formula (2.5) by $\exp(2\pi i \alpha F)$; F running through a system of positive integers; $F \leqslant n$. In view of the elementary formula:

$$\sum_{(a,g)=1} \exp 2\pi i \frac{a}{g} F = d\mu\left(\frac{g}{d}\right),$$

where $d = (q, F)$, we see that the solution quantity $r(n)$ can be expressed through the solution quantities of the congruences

$$F \equiv n \pmod{d},$$

where d/q and $F \leqslant n$ for each $q \equiv 0 \pmod{d}$; the numbers F must be supplied with the weights arising by the integration over the segments j_{ag}.

Hence, the initial additive problem for the number n in the case of the possibility of the levelling is reduced to a peculiar problem of the solution of the congruences for the moduli $d \leqslant \sqrt{N}$.

This reduction resembles the operation of the reduction of a curve with integer coefficients $f(x,y)=0$ for the prime moduli. In some cases some peculiar connections come to light, linking the operation described above with the well-known problem of Hasse about the behaviour of the Hasse function obtained as the product of the congruence zeta-functions arising by the reduction of a curve with respect to the prime moduli (the Hasse hypothesis, of M. Deuring, [5]). We shall see further that in the case of the additive problems with quadratic forms such an analogy is a rather close one.

We can remark that it is not known whether the levelling is possible for the binary Goldbach problem: $p_1 + p_2 = n$ (in the case of the twin primes problem: $p_1 - p_2 = 2$; $p_1 \leqslant n$ the levelling in a certain extended sense is

possible). Supposing that the levelling is possible one can effect a certain progress in the conditional solution of the binary Goldbach problem in using Riemann's extended hypothesis.

We can note also one more problem where the possibility of the levelling is essential. Consider a multilinear form

$F=\sum a_{j_1 j_2 \ldots j_k} x_{j_1} \ldots x_{j_k}$ of degree k with k^2 variables, for instance det (X) where X is a $k \times k$ matrix:

$$X = X_{kk} = \left\| \begin{matrix} x_{11} \ldots x_{1k} \\ x_{21} \ldots x_{2k} \\ \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ x_{k1} \ldots x_{kk} \end{matrix} \right\|.$$

We are interested in the solutions of the equation $F=n$ for which the point $(x_1, \ldots, x_{k^2})$ is situated in a given domain inside the cube $|x_i| < c\sqrt[k]{n}$, and, for $n \to \infty$ we consider the asymptotic of such solutions. This problem is of some interest for the analogies of the ergodic theorems for integer matrices. (cf. Ju. V. Linnik [13]).

We can fix in the proper domain k^2-2k conveniently chosen variables x_i and let the remaining ones change so that F is transformed into a quadratic form with very large coefficients (for instance, in the case when $F=\det(X)$, we can fix the first $n-2$ rows of the matrix X in the proper domain). In the equation $F=n$, F is transformed into a quadratic form with the coefficients increasing as $n^{1-2/k}$; the remaining variables must be situated in the prescribed domain. In case of a straightforward application of method (I) and of the well known estimates of Gauss sums to the equation $F=n$ the remainder term proves to have the same order as the principal one so that the levelling is necessary. By the levelling process some difficulties arise which do not allow to carry out the computation of the asymptotic to the end. We can formulate now only a weaker assertion about the property of being "everywhere dense".

THEOREM 1 (Ju. V. Linnik, B. F. Skubenko). *Let $n \to \infty$; consider the integer matrices*

$$X = X_{kk} = \left\| \begin{matrix} x_{11} \ldots x_{1k} \\ \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ x_{k1} \ldots x_{kk} \end{matrix} \right\|.$$

Consider the "fundamental cube" $|x_{ij}| \leqslant Cn^{1/k}$ ($C>1$ any constant, $\varepsilon>0$ a sufficiently small constant). If $n>n_0(\varepsilon, C)$; in each cube with the edges of length $\varepsilon n^{1/k}$ situated in the fundamental cube there is an integer point $(x_{11}, \ldots, x_{kk})$ such that for the corresponding matrix

$$X = \left\| \begin{matrix} x_{11} \ldots x_{1k} \\ \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \\ x_{k1} \ldots x_{kk} \end{matrix} \right\|$$

we shall have det $(X)=n$, *provided the obvious necessary conditions hold.*

For the case $k=2$ (2×2 matrices) the corresponding asymptotic can be found (cf. Ju. V. Linnik [12]). For the case $k=3$ the corresponding asymptotic was obtained recently by B. F. Skubenko and the author. It appears that in using the convenient method of levelling we can obtain the corresponding asymptotic (involving Haar's measure for the group of unimodular matrices) for the matrices of any order.

§3. The levelling in the additive problems involving quadratic foms. Analogies of Hasse's hypothesis for the Kloosterman sums

If an additive problem involves a positive quadratic form with two or more variables, the levelling process is possible. We get the problems of this type, for instance, considering the equations of the form $p+Q(x,y)=n$ (p, a prime; $Q(x,y)$, a positive binary quadratic form); $F(x_1,\ldots,x_k)=n$, where on the left-hand side there is a positive quadratic form. In the last case the levelling can be effected by means of the elementary theory of Hecke's operators. Let us explicate the method of such a levelling for the case of the equations:

$$x_1^2+\ldots+x_{2k}^2=n \quad (2k>4); \tag{3.1}$$

(this case is the most convenient one for the levelling process). The admissible error will be $O(n^{(\frac{1}{2}2k-\frac{1}{2})+\varepsilon})$. Let $P(x_1,\ldots,x_{2n})$ be any homogeneous polynomial of degree m with the integer coefficients.

Consider the problem of finding the asymptotic for the sum

$$S(n,P)=\sum_{x_1^2+\ldots+x_{2k}^2=n} P(x_1,\ldots,x_{2k}). \tag{3.2}$$

On the surface of the sphere $x_1^2+\ldots+x_{2k}^2=|\varrho|^2=1$ (ϱ being the radius vector), each polynomial $\pi(x_1,\ldots,x_{2k})$ can be represented as a sum of the homogeneous spherical polynomials of degree l, X_l (satisfying the Laplace equation $\Delta X_l=0$): $P=\sum_{l=0}^{h} X_l$ (cf. C. Pommerenke [18]).

Hence, under the condition $x_1^2+\ldots+x_{2k}^2=n$, we have;

$$P(x_1,\ldots,x_{2k})=\sum_{l=0}^{h}(\sqrt{n})^{m-l}X_l(x_1,\ldots,x_{2k}). \tag{3.3}$$

If $P(x_1,\ldots,x_{2k})$ is an inhomogeneous polynomial, we can break it up into homogeneous parts and we obtain the formula of the type (3.3) again. Now, the expression:

$$\vartheta(\tau,X_l)=\sum_{\substack{x_i=-\infty\\ i=1,2,\ldots,2k}}^{\infty} X_l(x_1,\ldots,x_{2k})\exp 2\pi i\tau(x_1^2+\ldots+x_{2k}^2) \tag{3.4}$$

is a modular form of the dimension $(u+l)$ in the upper half-plane $J\tau>0$. Hence the expression $S(n,P)$ in the formula (3,2) can be written as a linear combination with the coefficients $(\sqrt{n})^{m-l}$ of the coefficients of the term $\exp 2\pi i\tau n$ in the development of the modular forms (3.4).

For $l=0$ we obtain among them the principal θ function $\theta(\tau, X_0)$ corresponding to equation (3.1). It can be decomposed into the Eisenstein series and a Spitzenform; for $l>0$ we shall get only the Spitzenforms (cf. E. Hecke [8]). Thus the coefficient of $\exp(2\pi i\tau n)$ in the principal θ-function $\theta(\tau, X_0)$ is a linear expression of Eisenstein series, the Spitzenforms above mentioned and $S(n,P)$. Now, by the Hecke theory, we can express the coefficients of the said Spitzenforms through the finite basis of the eigenfunctions of the corresponding Hecke operators in the Spitzenform space. Now replacing the number n by the numbers $3n, 5n, 7n, 11n, \ldots, pn$ (p being prime) and using the properties of the eigenfunctions of the Hecke operators, we shall get a finite system of linear equations which is easily seen to give the expression of the solution number of equation (3.1) in terms of $S(np_i, P)$ only.

Hence, instead of investigating the asymptotic if the solutions of the equation (3.1) with the proper error term, we can look for the asymptotic of $S(n,P)$ with the proper error term. The polynomial $P(x_1, \ldots, x_{2k})$ can be chosen in this investigation in a rather arbitrary way to our convenience.

Now denoting by $r_4(m)$ the representation quantity of n by the sum of 4 squares we have by the well-known theorem of Jacobi:

$$r_4(m) = 8\sigma_1(m) \quad \text{if } m \text{ is uneven;}$$

$$r_4(m) = 24\sigma_1(m) \quad \text{if } m \text{ is even;}$$

here

$$\sigma_1(m) = \sum_{\substack{m=xy \\ x \text{ uneven}}} x$$

denotes the sum of the uneven devisors of m.

Let $P(x_1, \ldots, x_{2k})$ depend upon $x_{k-3}^2 + x_{k-2}^2 + x_{k-1}^2 + x_k^2$ only. Then $S(n, P)$ can be expressed by means of the sum:

$$S_1(n, P) = \sum_{\substack{x \text{ uneven} \\ x_1^2 + \ldots + x_{2k-4}^2 + xy = n}} xP(xy)$$

and an analogous sum involving the powers of 2. Consider the expression $\sum_{m=0}^{\infty} S(m, P) \exp(2\pi i\tau m)$. By what was said above, it can be expressed as a linear combination of the sums of the type $\sum_{n=0}^{\infty} S_1(n, P) \exp(2\pi i\tau n)$.

Put

$$\tau = \frac{i}{n} + \alpha, \quad \alpha \in (0.1).$$

Then:

$$\sum_{m=0}^{\infty} S_1(m,P) \exp(2\pi i\tau m) = \sum_{x_1, \ldots, x_{2k-3}} xP(xy) \exp 2\pi i\tau(x_1^2 + \ldots + x_{2k-4}^2 + xy). \quad (3.5)$$

The expression $\exp(2\pi i\tau xy)$ becomes very small for $xy \geqslant n \ln^2 n$. Now, if $xy \leqslant n \ln^2 n$ we have one or both of the two relations: $x \leqslant \sqrt{n} \ln n$ or $y \leqslant \sqrt{n} \ln n$. The sum (3.5) can be broken up into the first part, corresponding to the first relation and the second part corresponding to the second one. In the first part, $P(xy)$ being chosen conveniently, we can create the situation such that we can neglect with an admissible error the terms for which

(αx) is not very small so that only the terms with small (αx) will be essential. We can infer that for each x in (3.5) only the sum where

$$\alpha = \frac{a}{x} + \frac{\theta n^{\varepsilon_0}}{n};$$

$|\theta| \leqslant 1$, $\varepsilon_0 > 0$ arbitrarily small,will be of importance. If we form the integral $\int_0^1 (\ ) \exp(-2\pi i \alpha n) d\alpha$, where the expression (3.5) is to be put into the brackets, then the values of α for which

$$\alpha = \frac{a}{x} + \frac{\theta n^{\varepsilon_0}}{n} \quad (x \leqslant \sqrt{n} \ln n)$$

or the corresponding values for $y \leqslant \sqrt{n} \ln n$ will be of importance. Hence, in the problem (3.1) we can effect the levelling with the error $O(n^{(\frac{1}{4}2k-\frac{1}{2})+\varepsilon})$.

The levelling in the problem involving a general quadratic form with two or more variables is a little more sophisticated.

The possibility of levelling in the problem (3,1) with the error term $O(n^{(\frac{1}{4}2k-\frac{1}{2})+\varepsilon})$ leads to the fact that the problem of the optimal estimate of the error term in this problem is connected with the summation of the Kloosterman sums (2,1). The following hypothesis arises in this connection:

Hypothesis on Kloosterman's sums.

Let N be a large number;

$$T(N, g) = \sum_{x \bmod g} \exp \frac{2\pi i}{g} (x' + Nx) \quad \text{the Kloosterman sum,}$$

$$g_1 > N^{1/2-\varepsilon_0} \quad (\varepsilon_0 > 0 \text{ arbitrarily small}).$$

Then
$$\sum_{g \leqslant g_1} T(N, g) = O(g_1^{1+\varepsilon}) \quad \text{for each } \varepsilon > 0. \tag{3.6}$$

This hypothesis implies the estimate of the remainder term in the problem (2.2): $F(x_1, \ldots, x_k) = n$ for even k's: $|\sigma_F(n)| \gamma_F(\varepsilon) n^{(\frac{1}{4}k-\frac{1}{2})+\varepsilon}$. Moreover, the proof of this hypothesis would lead to considerable advances in the problem of counting the points inside the circle or in general an ellipsoid of even dimension.

This hypothesis can be considered as a certain analogy to the well known hypothesis of Hasse on the behaviour of congruence zeta-functions arising by the reduction of a given curve with respect to all prime moduli.

The estimate (3.6) is a theorem on the behaviour of the series $\sum_{g=1}^{\infty} T(N,g)/g^s$; the sums $T(N,g)$ appear in the process of reduction of a sequence of curves C_p: $y^p - y = Nx + 1/x$. (Cf. for instance L. Carlitz, S. Uchiyama [4]). At present we can indicate no way to the investigation of the hypothesis (3.6). However, in using certain properties relating essentially to the Hecke operators one can prove certain "one-sided" analogies of the said hypothesis (one-sided inequalities) relating to the summation of the Kloosterman sums.

§4. Some unequalities relating to the summation of the Kloosterman's sums

Let Q be a large prime; P, an uneven number such that $P \leqslant Q^{\varepsilon_0}$ (ε_0 a small constant); m, a positive number. Consider the Kloosterman sum:

$$S(P, m, Q) = \sum_{\substack{X \bmod Q \\ X \not\equiv 0 (\bmod Q)}} \exp \frac{2\pi i}{Q}(PX' + mX) \tag{4.1}$$

Let Q runs over the primes of the segment $[\overline{Q}, \overline{Q}+\Delta]$ where $\Delta = \overline{Q}^{1-\varepsilon_1}$ (ε_1 a small number chosen as a function of ε_0). Let

$$\sum_m = \sum_{(Q)} S(P, m, Q). \tag{4.2}$$

We form the sum:

$$\sum_{m \leqslant M} \Big|\sum_m\Big|^2 = T_M \tag{4.3}$$

Here M is a parameter to be fixed further. After some elementary transformations we get:

$$T_m = \sum_{m \leqslant M} \sum_{\substack{Q_1, Q_2 \\ Q_1 \neq Q_2}} \sum_{X \bmod Q_1 Q_2} \exp \frac{2\pi i}{Q_1 Q_2}(P(Q_1^2 + Q_2^2)X' + mX)$$
$$+ \sum_{m \leqslant M} |S(P, m, Q)|^2. \tag{4.4}$$

The last sum is estimated in using A. Weil's estimates (cf. for instance [19, 4]) as $O(M\overline{Q}^{2+\varepsilon})$ for any $\varepsilon > 0$. We have obviously $T_M \geqslant 0$. If we sum further over $M \leqslant M_1$, then over $M_1 \leqslant M_2$ and so on several times, we get the positive sums and the values of X such that $(X/Q_1 Q_2)$ is not very small will not be important and will contribute only an admissible error. Selecting the numbers $M, M_1, M_2, \ldots$ conveniently we can control the value $(X/Q_1 Q_2)$ for which it will be not possible to neglect the corresponding terms and so control the domain of summation over X (in all cases we shall make $X = O(\overline{Q}^{\varepsilon_2})$ ($\varepsilon_2 > 0$ small)

We can apply now the elementary relation:

$$XX' + Q_1 Q_2 (Q_1 Q_2)' \equiv 1 \pmod{XQ_1 Q_2} \tag{4.4}$$

the utility of which was pointed out to the author by I. V. Tshulanovski. By means of this relation we can transform the expressions of the type $\exp 2\pi i X'/Q_1 Q_2$ into the expressions of the type $\exp -2\pi i (Q_1 Q_2)'/X$ supplied with a complementary unessential factors.

We shall give the final result of the computation for a particular case.

THEOREM 2.

$$\sum_{x=1}^{\infty} y_r\left(\frac{P}{X}\right) X^r \varphi(X) \sum_{\substack{U \bmod X \\ (U, X)=1}} \cos \frac{2\pi P}{X}(U + U') \geqslant O(X^{r+2+\varepsilon}) \tag{4.6}$$

for each $\varepsilon > 0$.

Here $\varphi(x)$ is Euler's function, $r \geqslant 0$ any constant, and $y_r(z)$ is a function which can be expressed through Bessel functions. It changes very slowly for $0<z<1$ (and $0<c_1 y_r(z)<c_2$) in this interval); for $z>1, y_r(z)$ decreases very sharply so that roughly speaking, the summation in (4.6) is proceeding over $X \leqslant P$. The lower estimate (4.6) cannot be improved essentially. The proof of the corresponding upper estimate would imply the consequences similar to those of the hypothesis (3.6), in particular it would give the optimal estimate of the growth of the eigenvalues of Hecke operators for any integer dimension. The estimate (4.6) can be generalized considerably by taking other summation domains over X.

We get also certain unequalities connecting the sums over X of the Kloosterman sums involving the terms

$$\cos\frac{2\pi}{X}P_1(U+U'), \quad \cos\frac{2\pi P_2}{X}(U+U') \quad \text{and} \quad \cos\frac{2\pi}{X}(P_1 U+P_2 U').$$

§5. Some unequalities for the eigenvalues of the Hecke operators

The application of the levelling process in the case of theta-functions enables to infer certain new inequalities for the eigenvalues of the Hecke operators. Consider a theta-function of the type $(-k,1,1)$ (k being integer) (cf. E. Hecke [8]); let $\mathfrak{m}$ be the vector space of the Spitzenforms generated by the action of the Hecke operators on the Spitzenform arising from our theta-function by singling out the corresponding Eisenstein series. Let $F_1(\tau),\ldots,Fs(\tau)$ be modular eigenfunctions of the Hecke operators. Let $\lambda_\varrho(m)$ be an eigenvalue corresponding to the operator T_m. The well known hypothesis on the optimal estimates of the eigenvalues holds:

$$|\lambda_\varrho(m)| = O(m^{(\frac{1}{2}k-\frac{1}{2})+\varepsilon}) \quad \text{for each } \varepsilon>0. \text{ (See § 2.)}$$

It is well known that it is sufficient to prove this hypothesis for some powers of all prime numbers. Let p be a large prime and M a number such that $p \asymp M^{\varepsilon_0}$ ($\varepsilon_0>0$ a small constant). To prove the hypothesis mentioned above it is sufficient to prove that

$$\sum_{m\leqslant M} |\lambda_\varrho(pm)|^2 = O(M^k p^{k-1+\varepsilon}) \tag{5.1}$$

for any $\varepsilon>0$. This is a consequence of the multiplicative properties of the eigenvalues.

It is a regrettable fact that (5.1) up to now defies the attempts to prove it. However, one can prove the following estimate for the eigenvalues $\lambda_\varrho(n)$ mentioned above :

THEOREM 3. $\sum_{m\leqslant M} |\lambda_\varrho(pm+l)|^2 = O(M^k p^{k-1+\varepsilon})$ for $l=1,2,\ldots,p-1$. (5.2)

In these cases it is possible even to give an asymptotic expression of the left-hand side of (5.2) refining formula (5.2).

THEOREM 4. *If the conditions of the theorem* 3 *are fulfilled,*

$$\sum_{m\leq M} |\lambda_Q(pm+l)|^2 = \sigma_{pm} + O(M^k p^{k-\frac{3}{2}+\varepsilon}), \tag{5.3}$$

where σ_{PM} *does not depend upon* $l=1,2,\ldots,p-1$ *and has the estimate* $O(M^k p^{k-1+\varepsilon})$.

The case $l=0$ sharply differs from the other cases.

Let us indicate one more hypothesis which would imply the optimal estimate in the problem of the eigenvalues of the modular operators.

Let m be a given uneven number and the number runs over uneven numbers $\leqslant m$.

Hypothesis.

$$\sum_{\substack{Q\leqslant M\\ Q \text{ uneven}}} \sum_{\xi_Q^2\equiv 1 (\mathrm{mod}\, Q)} (-1)^{\frac{1}{2}Q-1} \exp\frac{2\pi i m \xi_Q}{Q} = O(m^{1/2+\varepsilon}). \tag{5.4}$$

Here ξ_Q runs over all the solutions of the congruence $\xi_Q^2 \equiv 1 \pmod{Q}$. (For instance, for a prime Q, $\xi_Q \equiv \pm 1 \pmod{Q}$). One can prove that the left-hand side of (5.4) has the estimate $O(m^{\frac{3}{4}+\varepsilon})$.

§6. Some asymptotic analogies of the Hecke operators. A problem on multiple Gauss sums

The theory of the Hecke operators enables to express the quantity of the solutions of the equations $F(x_1,\ldots,x_{2k})=p$ by means of a "discrete" operator:

$$\vartheta \,|\, T_p = \varepsilon(p) p^{k-1} \vartheta(p\tau) + \frac{1}{p}\sum_{l=0}^{P-1} \vartheta\left(\frac{\tau+l}{p}\right) \tag{6.1}$$

as well as by means of the integral operator indicated in 1), § 1.

The Hecke operators act only for the space of the modular forms of integer dimension, in particular for the theta-functions generated by positive quadratic forms with an even number of variables.

We shall consider here the action of essentially the same operators in a different domain. Our considerations will be asymptotic ones.

Let p be a large prime, N a big parameter; $\theta_1 \asymp 1/N$; $s_j>0$ fixed integer numbers; we consider the product of k theta-functions ($k\geqslant 1$ being even or uneven).

$$\prod = \prod_{j=1}^{k} \sum_{\xi_j=-\infty}^{\infty} \exp\left(-\frac{1}{N} - \frac{2\pi i A}{Q} - i\theta_1\right) s_j \xi_j^2. \tag{6.2}$$

Here A and Q are integer numbers; $(A,\mathrm{C})=1$; Q increases together with N (we shall fix further in which way)

Apply to Π the operator τ_l corresponding to

$$\frac{1}{p}\vartheta\left(\frac{\tau+l}{p}\right)$$

in formula (6.1) thus obtaining

$$\tau_l(\prod)=\frac{1}{p}\prod_{j=1}^{k}\sum_{\xi_j=-\infty}^{\infty}\exp\left(-\frac{1}{Np}-\frac{2\pi i A_l}{Qp}-i\theta\right)s_j\xi_j^2, \tag{6.3}$$

where $A_l=A+lQ$; $\theta=\theta_1/p$.

Applying the elementary formula of the transformations of θ-functions, we find:

$$\tau_l(\prod)=\frac{1}{Qp^{n+1}}\left(\frac{\pi_k}{s_1s_2\dots s_k}\right)^{1/2}\prod_{j=1}^{k}\left(\frac{1}{Np}-i\theta\right)^{-1/2}.$$

$$S_{s_1A_l,\,Q_p}\dots S_{s_kA_l,\,Q_p}\sum_{\nu_1,\dots,\nu_k=-\infty}^{\infty}x_{\nu_1}\dots x_{\nu_k}\exp-\frac{2\pi i}{Qp}4'A'_e.$$

$$(s_1'\nu_1^2+\dots+s_n'\nu_n^2)\exp-\frac{\pi^2}{(Qp)^2\left(\frac{1}{Np}-i\theta\right)}(s_1^{-1}\nu_1^2+\dots+s_k^{-1}\nu_k^2, \tag{6.4}$$

where $x_{2_i}=2(\nu_i\neq 0)$; $x_0=1$, $S_{s_jA_l,\,Q_p}$ Gauss sum, the sign′ is the reciprocal element mod Q_p. We shall suppose Q to be uneven; $Q\asymp\sqrt{N}$, $Q_p\asymp\sqrt{N}p$. We have:

$$S_{s_jA_l,\,Q_p}=\left(\frac{S_jA_l}{Qp}\right)(Qp)^{1/2}\quad\text{if}\quad(A_l,Q_p)=1. \tag{6.5}$$

We introduce now the operator $T^{(p)}$ corresponding to

$$\frac{1}{p}\sum_{l=0}^{p-1}\vartheta\left(\frac{\tau+l}{p}\right).$$

We shall sum over $l=0,1,2,\dots,p-1$ taking into account (6.5). If $(A_l,Q_p)=1$ and k is even, we get

$$\prod_{j=1}^{k}\left(\frac{S_jA_l}{Qp}\right)=\prod_{j=1}^{k}\left(\frac{S_j}{Qp}\right),$$

and if we drop the only term for which $(A_l,Qp)>1$, we can replace $T^{(p)}(\Pi)$ with an admissible error through:

$$T_1^{(p)}(\prod)=\frac{1}{Q^{k/2}p^{k/2+1}}\left(\frac{\pi^k}{s_1s_2,\dots,s_k}\right)^{1/2}\left(\frac{1}{Np}-i\theta\right)^{-k/2}.$$

$$\sum_{\nu_1,\dots,\nu_n=-\infty}^{\infty}\varkappa_{\nu_1},\dots,\varkappa_{\nu_k}\cdot\sum_{\substack{l=0\\A_l\not\equiv 0(\mathrm{mod}\,p)}}^{p-1}\exp-\frac{2ni}{Qp}4A_l(s_1^1\nu_1^2+\dots+s_k^1\nu_k^2).$$

$$\exp-\frac{\pi^2}{(Qp)^2\left(\frac{1}{Np}-i\theta\right)}(s_1^{-1}\nu_1^2+\dots+s_n^{-1}\nu_k^2). \tag{6.6}$$

In connection with the summing over A_l, the values of $\Phi(\nu_1,\dots,\nu_n)$ adjoint to $s_1\nu_1^2+\dots+s_n\nu_n^2$ are singled out, which are divisible by p; the

corresponding error is an admissible one. Each of these values up is supplied with a coefficient damping them as e^{-cr}. This enables to express the solution quantity of the equation $\Phi(\nu_1,\ldots,\nu_k)=p$ through the linear combinations of the expressions $\tau^{(p)}(\Pi)$. For the described case (n even, $s_j>O$) it would be simpler and more convenient to use Hecke operators. But the zone of action of the operators $\tau^{(p)}$ is somewhat larger.

Consider the case of the uneven k (for instance, $k=3$, the case of ternary forms). We shall suppose here that the number Q is divisible by p and contains p in an uneven power: $p^{2m+1}//Q$, In such case, for $A_l\not\equiv 0(\mathrm{mod}\, p)$ we shall get:

$$\prod_{j=1}^{k}\left(\frac{s_jA_l}{Qp}\right)=\prod_{j=1}^{k}\left(\frac{s_jA}{Q_1}\right),$$

so that this number does not depend upon l. This enables to express with an admissible error the solution quantity of the equation $\Phi(\nu_1,\ldots,\nu_k)=p$ also for the case of an uneven k by means of a linear combination of the operators $\tau^{(p)}$; however, in the argument

$$-\frac{1}{N}+\frac{2\pi iA}{Q}-i\theta$$

(see (6.2)), the number Q must contain p in an uneven power. Hence, in the case of theta-functions of a non-integer dimension, we can express the coefficients by means of the "discrete" operators acting on the products of the theta-functions the arguments of which can change not in the whole half-plane but in the points the imaginary components of which obey a certain simple arithmetic condition, connected with the number p to be represented.

We can remark also that the same considerations are valid for a certain degree for the weighted representations of the number p by indefinite quadratic forms. For this we must take in formula (6.2) the numbers A/Q with positive as well as negative signs. For the forms with an even number of variables we get the expressions of the weighted representation number with the admissible error through the operators $\tau^{(p)}$, for the forms with the uneven number of variables one must impose certain requirements of an arithmetical nature of the type described above on the arguments of the product of theta-functions.

The same arithmetical conditions enable us to obtain somewhat more general operators with the analogous properties: $T_\xi^{(p)}$ defined as

$$\frac{1}{p}\sum_{l=0}^{p-1}\exp 2\pi i\frac{\xi l}{p}T_l;$$

for $\xi=0$ they coincide with the former ones.

We turn now to problem of multiple Gaussian sums; it is connected with the Hecke operators and the problem of the estimates of their eigenvalues.

Let $f(\tau)=f_1(\tau)-f_2(\tau)$ be the difference of two theta-functions which forms a Spitzenform of an integer dimension $-k$.

For instance, say, $f(\tau)=\sum\exp 2\pi iF_1\tau-\sum\exp 2\pi iF_2\tau$, where F_1 and F_2 are positive quadratic forms of the same discriminant and genus.

Let
$$\tau = \frac{i}{N} + 2\pi\frac{A}{Q} + \theta$$

be a point of "general position". This means that A/Q is an irreducible fraction and Q does not differ essentially from $\sqrt{N}$, say:

$$\sqrt{\frac{N}{\ln N}} \leqslant Q \leqslant \sqrt{N \ln N} \quad \text{and} \quad \theta = O\left(\frac{1}{Q^2}\right).$$

Let p be a large prime and T_p the Hecke operator for p so that:

$$f \mid T_p = \varepsilon(p) p^{k-1} f(p\tau) + \frac{1}{p}\sum_{l=0}^{p-1} f\left(\frac{\tau+l}{p}\right) = \sum_{\varrho} \lambda_\varrho(p) f_\varrho(\tau),$$

where $\lambda_\varrho(p)$ and $f_\varrho(\tau)$ are the corresponding eigenvalues and eigenfunctions. The admissible error here will be considered $O(p^{(\frac{1}{2}k-\frac{1}{2})+\varepsilon})$. We suppose that $p \asymp N^{\varepsilon_0}$, where $\varepsilon_0 > O$ is a small constant. It is easy to see that the first term in the above formula: $\varepsilon(p)p^{k-1}f(p\tau)$ can be deleted with an admissible error. Thus the problem of the estimate of the eigenvalues is reduced to the problem of proving in a conveniently chosen point of general position:

$$\tau = \frac{i}{N} + 2\pi\frac{A}{Q} + \theta,$$

the estimate:
$$\sum_{l=0}^{p-1} f\left(\frac{\tau+l}{p}\right) = O(N^k p^{k+\frac{1}{2}+\varepsilon})$$

while the trivial estimate will be $O(N^k p^{\mu+1+\varepsilon})$. The characteristic feature of this problem is that we are allowed to select the point of general position arbitrarily, to our convenience (say, to take $Q = q^s$ where q is a prime, for instance the number p itself), and $\theta = 0$.

For the case where $f(\tau)$ is generated by the forms of one and the same genus: $(F_1 = F_1(x_1, \ldots, x_{2k}), F_2 = F_2(x_1, \ldots, x_{2n})$ we have the following problem on multiple Gaussian sums:

Consider the difference of the type:

$$\sum_{l=0}^{p-1} \sum_{x_1, \ldots, x_{2k}} \exp\left(-\frac{1}{Np} + 2\pi i\left(\frac{A}{Q} + \frac{l}{p}\right) F_1(x_1, \ldots, x_{2k})\right.$$

$$-\sum_{l=0}^{p-1} \sum_{x_1, \ldots, x_{2n}} \exp\left(-\frac{1}{Np} + 2\pi i\left(\frac{A}{Q} + \frac{l}{p}\right) F_2(x_2, \ldots, x_{2k})\right. \tag{6.7}$$

To study that we can choose a number A/Q as an irreducible fraction such that $\sqrt{N}/\ln N \leqslant Q \leqslant \sqrt{N \ln N}$ and that the difference (6.7) will be estimated as $O(N^k p^{k+\frac{1}{2}+\varepsilon})$.

Apparently it is worth while to choose Q of the form $Q = p^s$ and then to select the corresponding number A.

REFERENCES

[1]. ANDRIANOV, A. N., Generalization of a theorem of M. Eichler from the theory of quaternary quadratic forms. *Dokl. Akad. Nauk S.S.S.R.*, 141, 1 (1961), 9–12. (In Russian.)

[2]. —— On the analytic arithmetic of quadratic forms with an uneven number of variables, in connection with the theory of modular forms. *Dokl. Akad. Nauk S.S.S.R.*, 144, 4 (1962). (In Russian.)

[3]. BREDIKIN, B. M., Binary additive problems with the prime numbers. *Dokl. Akad. Nauk S.S.S.R.*, 142 (1962), 766–768. (In Russian.)

[4]. CARLITZ, L. & UCHIYAMA, S., Bounds for exponential sums. *Duke Math. J.*, 24, 1 (1957), 37–41.

[5]. DEURING, M,, The zeta-functions of algebraic curves and varieties. *Report of Internat. Colloq. on Zeta-functions.* (Bombay, 1956), 89–101.

[6]. EICHLER, M., Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion. *Arch. Math.*, 5 (1954), 355–366.

[7]. —— Quadratische Formen und Modulfunktionen. *Acta Arith.*, 4 (1958), 217–239.

[8]. HECKE, E., *Analytische Arithmetik der positiven Quadratischen Formen.* København, 1940.

[9]. KLOOSTERMAN, H. D., On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$. *Acta Math.*, 49 (1926), 407–464.

[10]. LINNIK, JU. V., Some conditional theorems concerning binary Goldbach's problem. *Izv. Akad. Nauk S.S.S.R., ser. mat.*, 16 (1952), 503–520. (In Russian.)

[11]. —— Adding prime numbers to the powers of one and the same number. *Mat. Sb. N.S.* 32, 1 (1953), 3–60. (In Russian.)

[12]. —— Asymptotic distribution of reduced binary quadratic forms in connection with the Lobatshevskian geometry. *Vestnik Leningradskogo Universiteta*, 2 (1955), 3–32. (In Russian.)

[13]. —— Application of non-euclidean geometries to the theory of Dirichlet's characters. Analogue of ergodic theorems. *Trudi III Vsesoius. Matem. Sjesda*, III (1958), 21–29. (In Russian.)

[14]. —— The variance method in binary additive problems. *Isdat. Len. Gos. Universiteta*, 1961. (In Russian.)

[15]. MALYSHEV, A. V., On the distribution of integer points upon a 4-dimensional sphere. *Dokl. Akad. Nauk S.S.S.R.*, 114, I (1958), 25–28. (In Russian.)

[16] —— Asymptotic distribution of the integer points on some ellipsoids. *Izv. Akad. Nauk S.S.S.R., ser mat.*, 21, 4 (1957), 457–500. (In Russian.)

[17]. —— On the representation of integer numbers by positive quadratic forms with 4 or more variables. *Dokl. Akad. Nauk S.S.S.R.*, 133, 6 (1960), 1294–1297. (In Russian.)

[18]. POMMERENKE, CH., Über die Gleichverteilung von Gitterpunkten auf m-dimensionalen Ellipsoiden. *Acta Arith.*, 5 (1959), 227–257.

[19]. WEIL, A., The number of solutions of the equations in finite fields. *Bull. Amer. Math. Math. Soc.*, 55 (1949), 497–508.

[20]. VINOGRADOV, A. I., On an "almost binary" problem. *Izv. Akad. Nauk S.S.S.R., ser. mat.*, 20, 6 (1956), 713–750. (In Russian.)

[21]. —— Generalization of the formula of Kloosterman. *Dokl. Akad. Nauk S.S.S.R.*, 147, 5 (1962). (In Russian.)

A CLASS OF DOUBLY TRANSITIVE PERMUTATION GROUPS

By MICHIO SUZUKI

I will consider here a class of doubly transitive permutation groups which have been the object of recent investigations. The problem of determining all multiple transitive permutation groups is an old one and it still awaits a solution. I do not consider this problem at all; rather I consider a particular class of groups.

It has been known that certain classes of permutation groups are characterized by their group theoretical properties. I will mention a few results along this line. I use the following notation. A permutation group on a set Ω is usually denoted by G. I consider only a transitive permutation group on a finite set. If $a, b, \ldots$ are distinct elements of Ω, I denote by $G_{a,b,\ldots}$ the subgroup of G consisting of permutations leaving each of these elements $a, b, \ldots$.

A Frobenius group is a transitive permutation group satisfying $G_{a,b,} = \{1\}$ for any pair a, b of elements of Ω. The class of Frobenius groups is characterized by several group theoretical properties. If H is the subgroup G_a of a Frobenius group G, H satisfies the property

$$H \cap g^{-1}Hg = \{1\} \quad \text{for all} \quad g \in G - H.$$

This property characterizes the class of Frobenius groups: namely a group G containing a subgroup H with the above property is a Frobenius group. A famous theorem of Frobenius asserts that in a Frobenius group G the totality of elements which moves every element of Ω forms together with the identity a normal subgroup N which is called the Frobenius kernel of G and satisfies the property that

$$C_G(x) \subseteq N \quad \text{for} \quad x \neq 1, x \in N.$$

Here $C_G(x)$ denotes the centralizer of x in G. Again this is a characteristic property of the class of Frobenius groups. The structure of Frobenius groups is fairly well known. The factor group G/N by the Frobenius kernel has been studied by Zassenhaus who classified the possible structure of G/N, and the structure of N itself is nilpotent by a theorem of Thompson.

Zassenhaus has studied a class of triply transitive groups G satisfying $G_{a,b,c} = \{1\}$. This is a class of triply transitive groups of smallest possible order, and the result of Zassenhaus is that groups in this class are the groups of all the linear fractional transformations of one variable over some near fields, whose structures are completely determined.

I considered a little wider class of groups which are called (Z)-groups. A (Z)-group is a doubly transitive permutation group G satisfying the following two conditions:

(1) $G_{a,b,c} = \{1\}$ and

(2) There is no regular normal subgroup.

The second restriction is not essential. In fact if a doubly transitive permutation group satisfies (1) but not (2) then the group is a group of semi-linear transformations

$$x' = ax^{\sigma} + b$$

over a finite near field. The structure of (Z)-groups is studied by Feit, Itô and myself. It is convenient to denote the degree of G by $1+g$. Then the group G_a is a Frobenius group on q elements. Hence G_a contains a normal subgroup Q of order exactly q. By using a theory of characters Feit proved that for a (Z)-group q must be a power of a prime number. The case when this prime number p is greater than 2 is settled by Ito, again by using a theory of characters. The result in this case is that a (Z)-group G on a set Ω is a subgroup of index $\leqslant 2$ of a triply transitive group on Ω which satisfies the condition (1). Hence again G is a group of linear fractional transformations over some near fields. The case when $p=2$ caused a difficulty. I call a (Z)-group with $p=2$ a (ZT)-group. It is known that the totality of linear fractional transformations over a field of characteristic 2 is a (ZT)-group, when it is considered as a permutation group on points of the projective line over the field. Recently I discovered a class of new simple groups. They are defined as a subgroup of $GL(4,q)$ where q is a power of 2 and is non square. Originally they are given as a group generated by a certain set of 4×4 matrices. Perhaps it is easy to define them as a set of fixed elements by a special automorphism of the classical group $Sp(2)$, which in a sense exchanges two dots of Dynkin Diagram of B_2. This new definition is due to Ree and Steinberg. One property of this class of groups is that every member is a (ZT)-group. Tits remarked that in the three dimensional projective space over $GF(q)$ there is a certain oval consisting of $1+q^2$ points and the group in question is defined as the totality of projective transformations leaving this oval invariant. The permutation representation on points of this oval is doubly transitive and satisfies the properties (1) and (2) if $q>2$. It was rather difficult to prove that there are no other (ZT)-groups.

A class of (ZT)-groups is characterized by some group theoretical properties. Namely one has the following theorem. A group G is a (ZT)-group if and only if G is a non-abelian simple group containing a subgroup H such that

(i) the center of H has even order, and
(ii) $C_G^*(x) \subseteq H$ for $x \neq 1, x \in H$.

Here $C_G^*(x)$ denotes the generalized centralizer: i.e.

$$C_G^*(x) = \{y \mid y^{-1}xy = x \text{ or } x^{-1}\}.$$

The conditions (i) and (ii) can be stated in weaker forms. For example (ii) is required only for involutions in H and for other elements the conditions

$$C_G^*(x) \subseteq N_G(H)$$

are sufficient.

A (ZT)-group satisfies other interesting group theoretical properties. In a (ZT)-group G the centralizer of every non-identity element is nilpotent.

This property is of course shared by other groups. However it can be proved that if a non-abelian simple group G satisfies the above property, then G is either a (ZT)-group or one of the linear groups.

The simple groups in the first class discovered by Ree also admit a doubly transitive representation. Ree's groups do not satisfy $G_{a,b,c}=\{1\}$ for some triple a,b,c but the group $G_{a,b,c}$ is very small in any case. Both Ree's groups and (Z)-groups have the following property in common:

(*) The group G_a contains a normal subgroup Q, which is transitive and regular on $\Omega-\{a\}$.

I studied a class of doubly transitive permutation group G of degree n satisfying the condition (*). The main result is the following theorem.

Let G be as above. If moreover the degree n is odd and if the group G_a/Q is a solvable group of odd order, then G is isomorphic to one of the following three groups: (i) *an extension of a (ZT)-group,* (ii) *an extension of a unitary group of dimension* 3 *over a field of characteristic* 2 *or* (iii) *an extension of the group of all linear transformations over a near field.*

Ree's groups do not come in because I assumed that the degree is odd. In all cases extensions are obtained by applying automorphisms of the underlying (near) field.

The proof of the above theorem requires many tools, including the theory of characters and of transfers. A recent theorem of G. Higman on the structure of a particular class of 2-groups is used to determine the structure of Sylow 2-groups of G. The ideas in proof are quite similar to the ones used to determine the structure of (ZT)-groups. The last case (iii) is a sort of a degenerated case. One has the case (iii) if and only if the group Q contains only one involution.

The assumption (*) seems quite essential in this study. Also the assumption of solvability of G_a/Q is used to determine the structure of $G_{a,b}$. Although my method does not apply to the groups of even degree, there are several indications that one might be able to classify all groups with the condition (*).

DUALITY THEOREMS IN GALOIS COHOMOLOGY OVER NUMBER FIELDS

By JOHN TATE

1. Notation and terminology

Let X be a Dedekind ring with field of fractions k and let C be a commutative group scheme over X. Except in the special case $X=\mathbf{R}$ or $\mathbf{C}$ (real or complex field) we put, for all $r\in\mathbf{Z}$,

$$H^r(X, C)=\varinjlim_{K} H^r(G_{K/k}, C_Y),$$

the direct limit taken over all finite Galois extensions K of k in which the integral closure Y of X is unramified over X, where $G_{K/k}$ denotes the Galois group of such an extension, and where C_Y denotes the group of points of C with coordinates in Y. For example, if $X=k$, our notation coincides with that of [10]. For any X, the group $H^r(X,C)$ is the r-th cohomology group of the profinite group $G_X=\varprojlim_{K} G_{K/k}$ (fundamental group of Spec X) with coefficients in the G_X-module $\varinjlim C_Y$ of points of C with coordinates in the maximal unramified extension of X; a general discussion of the cohomology theory of profinite groups can be found in [5]. In the special case $X\approx\mathbf{R}$ or $\mathbf{C}$ we put

$$H^r(\mathbf{R}, C)=\hat{H}^r(G_{\mathbf{C}/\mathbf{R}}, C) \quad \text{and} \quad H^r(\mathbf{C}, C)=\hat{H}^r(G_{\mathbf{C}/\mathbf{C}}, C)=0,$$

where $\hat{H}$ denotes the complete cohomology sequence of the finite group G, in general non trivial in negative dimensions ([2], Ch. 12).

In our applications, X will be a ring associated with an algebraic number field, or with an algebraic function field in one variable over a finite constant field, and the group scheme C will be one of two special types, which we will denote by M and A, respectively. By M we shall always understand (the group scheme of relative dimension zero over X associated with) a finite G_X-module whose order, $|M|=\text{card } M$, is prime to the characteristics of the residue class fields of X. By A we shall denote an abelian scheme over X (i.e., an abelian variety defined over k having "non-degenerate reduction" at every prime of X).

Underlying our whole theory is the cohomology of the multiplicative group, $\mathbf{G}_m$, as determined by class field theory. For any M, we put $M'=\text{Hom}\,(M,\mathbf{G}_m)$. By our assumption that $|M|$ is invertible in X, we see that M' is a group scheme of the same type as M, namely the one associated with the G_X-module $M'=\text{Hom}(M,\mu)$ where μ denotes the group of roots of unity. Moreover we have $|M|=|M'|$, and $M\approx(M')'$. For any A we put $A'=\text{Ext}(A,\mathbf{G}_m)$, the dual abelian variety; then $A\approx(A')'$ by "biduality". Our aim is to discuss dualities between the cohomology of C and C' in both cases, $C=M$ and $C=A$. Notice that $\text{Ext}(M,\mathbf{G}_m)=0$ and $\text{Hom}(A,\mathbf{G}_m)=0$, so that in each case, C' denotes the only non-vanishing group in the sequence $\text{Ext}^r(C,\mathbf{G}_m)$. Thus our results are presumably special cases of a vastly more

general hyperduality theorem for commutative algebraic groups envisaged by Grothendieck, involving all the $\mathrm{Ext}^r(C, \mathbf{G}_m)$ simultaneously.

Finally, for any locally compact abelian group H, we let H^* denote its Pontrjagin character group.

2. Local results

Let k be either $\mathbf{R}$ or $\mathbf{C}$ (archimedean cases), or be complete with respect to a discrete valuation with finite residue class field (non-archimedean case). By local class field theory we have a canonical injection $H^2(k, \mathbf{G}_m) \to \mathbf{Q}/\mathbf{Z}$ which associates with each element of the Brauer group of k its "invariant" which is a rational number (mod. 1). Hence the cup product with respect to the canonical pairing $M \times M' \to \mathbf{G}_m$ gives pairings $H^r(k, M) \times H^{2-r}(k, M') \to \mathbf{Q}/\mathbf{Z}$.

THEOREM 2.1. *For all M and r the group $H^r(k, M)$ is finite and the pairing just discussed yields an isomorphism $H^r(k, M) \approx H^{2-r}(k, M')^*$.*

In particular we have $H^r(k, M) = 0$ for $r > 2$ when k is non-archimedean. In fact, even more is true in that case, namely the group G_k has strict cohomological dimension 2 in the sense of [5]. This fact, together with the theorem, can be proved easily using results of Nakayama [11], by writing $M = F/R$, where F is a $\mathbf{Z}$-free G_k-module.

THEOREM 2.2. *If k is archimedean, then $H^r(k, M)$ is an elementary abelian 2-group whose order is independent of r. If k is non-archimedean and $\mathfrak{o}$ is the valuation ring in k, then the "Euler-characteristic" of M has the value*

$$\chi(k, M) = \frac{|H^0(k, M)|\,|H^2(k, M)|}{|H^1(k, M)|} = \frac{1}{(\mathfrak{o} : |M|\mathfrak{o})}.$$

The archimedean case is trivial. In the non-archimedean case one uses the multiplicativity of χ to reduce to various special types of simple M's, only one of which is difficult. In that one the determination of χ is essentially equivalent to a result of Iwasawa [7] on the structure of $K^*/(K^*)^p$ as $G_{K/k}$-module, for a certain type of extension K/k.

In case of an abelian variety A defined over k the relationship $A' = \mathrm{Ext}(A, \mathbf{G}_m)$ leads to a "derived cup product" pairing:

$$H^r(k, A) \times H^{1-r}(k, A') \to \mathbf{Q}/\mathbf{Z},$$

as explained in [16]. The group $H^0(k, A)$, which is the group of rational points of A in k (modulo its connected component in case $k = \mathbf{R}$ or $\mathbf{C}$) is compact and totally disconnected because A is complete and k locally compact. On the other hand, we view $H^1(k, A)$ as discrete and have then

THEOREM 2.3. *For all A and r, the pairing just mentioned yields an isomorphism $H^r(k, A) \approx H^{1-r}(k, A')^*$ (possibly provided we ignore the p-primary components of the groups in case k is of characteristic $p > 0$).*

In the archimedean case this theorem is due to Witt [17]. For k non-archimedean of characteristic 0, the case $r = 0$ and 1 is proved in [16]. One

can simplify that proof and at the same time extend the result to all r (i.e., show $H^2(k,A)=0$ in the non-archimedean case), by applying theorems 2.1 and 2.2 to the kernel, M, of the isogeny $A \xrightarrow{m} A$. The same method works for k of positive characteristic, with the proviso in the theorem. Although that proviso is probably unnecessary, new methods will be required to remove it, possibly those of Shatz [15], where the analog of theorem 2.1 is proved for arbitrary finite commutative group schemes M over k.

Suppose now that k is non-archimedean with valuation ring $\mathfrak{o}$. Let M be a finite $G_{\mathfrak{o}}$-module such that $|M|$ is invertible in $\mathfrak{o}$.

THEOREM 2.4. *We have* $|H^0(\mathfrak{o},M)|=|H^1(\mathfrak{o},M)|$ *and* $H^r(\mathfrak{o},M)=0$ for $r>1$. *In the duality of Theorem* 2.1 *between* $H^1(k,M)$ *and* $H^1(k,M')$, *the subgroups* $H^1(\mathfrak{o},M)$ *and* $H^1(\mathfrak{o},M')$ *are the exact annihilators of each other.*

The first statements follow from the fact that $G_{\mathfrak{o}}$ is a free profinite group on one generator. The annihilation results from $H^2(\mathfrak{o},\mathfrak{G}_m)=0$. The *exact* annihilation now follows by counting, using theorems 2.1 and 2.2.

THEOREM 2.5. *If A is an abelian scheme over* $\mathfrak{o}$, *we have* $H^1(\mathfrak{o},A)=0$.

To prove this one has only to combine results of Lang [8] and Greenberg [6].

3. Global results

Let k be a finite extension of $\mathbf{Q}$ (case (N)), or a function field in one variable over a finite field (case (F)). Let S be a non-empty (possibly infinite) set of prime divisors of k, including the archimedean ones in case (N), and let k_S denote the ring of elements in k which are integers at all primes P *not* in S. For example, if S is the set of all prime divisors of k, then $k_S=k$. Let M be a finite module for the Galois group of the maximal extension of k unramified outside S, and such that $|M|k_S=k_S$. For each prime P *in* S, let k_P denote the completion of k at P. The localization maps $H^r(k_S,M)\to H^r(k_P,M)$ taken all together yield a map

$$H^r(k_S,M)\xrightarrow{\alpha_r}\prod_{P\in S}{}' H^r(k_P,M),$$

where the symbol $\prod'$ denotes the (compact) *direct product* for $r=0$, the (locally compact) *restricted direct product relative to the subgroups* $H^1(\mathfrak{o}_p,M)$ for $r=1$, and the (discrete) *direct sum* for $r\geqslant 2$. By Theorems 2.1 and 2.4, our local dualities yield isomorphisms

$$\prod_{P\in S}{}' H^r(k_P,M)\approx\Big[\prod_{P\in S}{}' H^{2-r}(k_P,M')\Big]^*.$$

Thus by duality we obtain maps

$$\prod_{P\in S}{}' H^r(k_P,M)\xrightarrow{\beta_r}H^{2-r}(k_P,M')^*,$$

namely $\beta_r=(\alpha'_{2-r})^*$, where α' is to M' as α is to M.

Let $\mathrm{Ker}^r(k_S,M)$ denote the kernel of α_r, that is, the group of elements in

$H^r(k_S, M)$ which are zero locally at all primes $P \in S$. There is a canonical pairing

$$(*) \qquad \mathrm{Ker}^2(k_S, M) \times \mathrm{Ker}^1(k_S, M') \to \mathbf{Q}/\mathbf{Z}$$

defined as follows: we represent the cohomology classes to be paired by a 2-cocycle f and a 1-cocycle f'. Then for each $P \in S$ we have, over k_P, a 1-cochain g_P and a 0-cochain g'_P such that $f = \delta g_P$ and $f' = \delta g'_P$. Also, since $H^3(k_S, \mathbf{G}_m)$ has no non-zero elements of order dividing $|M|$, there is, over k_S, a 2-cochain h with coefficients in $\mathbf{G}_m$, such that $f \cup f' = \delta h$. Then, over k_P, we have $\delta(g_P \cup f') = \delta h = \delta(f \cup g'_P)$ and $\delta(g_P \cup g'_P) = f \cup g'_P - g_P \cup f'$, so that for each P the cochains $(g_P \cup f') - h'_P$ and $(f \cup g'_P) - h$ are cocycles representing the same class, say x_P, in $H^2(k_P, \mathbf{G}_m)$. We pair our original elements to the sum (over $P \in S$) of the invariants of these x_P; it is easy to see that the result is independent of the choices involved.

THEOREM 3.1. (*a*) *The pairing* (*) *just discussed is a perfect duality of finite groups.*

(*b*) α_0 *is injective,* β_2 *is surjective, and for* $r = 0, 1, 2$ *we have* $\mathrm{Im}\ \alpha_r = \mathrm{Ker}\ \beta_r$.
(*c*) α_r *is bijective for* $r \geqslant 3$.

Notice that these statements imply, and are, in turn, summarized by, the existence of an exact sequence:

$$\begin{tikzcd}
0 \arrow[r] & H^0(k_S, M \arrow[r, "\alpha_0"] & \prod_{P \in S} H^0(k'_P, M) \arrow[r, "\beta_0"] & H^2(k_S, M')^* \arrow[r] & H^1(k_S, M) \arrow[dr, "\alpha_1"] & & \\
& & & & & \prod_{P \in S} H^1(k_P, M) \arrow[dl, "\beta_1"] & \\
0 & H^0(k_S, M')^* \arrow[l] & \prod_{P \in S} H^2(k_P, M) \arrow[l, "\beta_2"'] & H^2(k_S, M) \arrow[l, "\alpha_2"'] & H^1(k_S, M')^* \arrow[l] & &
\end{tikzcd}$$

together with isomorphisms

$$H^r(k_S, M) \simeq \prod_{P \text{ real}} H^r(k_P, M) \quad \text{for} \quad r \geqslant 3,$$

where the unlabeled arrows in the exact sequence require the non-degeneracy of the pairing (*) for their definition.

I understand that a large part of Theorem 3.1 has been obtained independently by Poitou, and I suspect that the theorem is closely related to results of Shafaryevitch on the extension problem to which he alluded in his talk at this Congress.

If $M = \mu_m$, the group of mth roots of unity, then Theorem 3.1 summarizes well-known statements in class field theory. For general M, all statements of the theorem except case $r = 1$ of (*b*) can be proved by considering the pairing $M \times \mathrm{Hom}(M, C) \to C$, where C is the S-idele-class group of the maximal extension of k unramified outside S; denoting by G the Galois group of that extension, one shows that the resulting pairing $H^2(k_S, M) \times \mathrm{Hom}_G(M, C) \to \mathbf{Q}/\mathbf{Z}$ is non-degenerate, except that in case (N), there is a kernel on the right-hand side, namely the norm from K to k of $\mathrm{Hom}_{\mathbf{Z}}(M, D_K)$, where D_K is the connected component of the idele class group of a sufficiently large finite extension K of k. For finite S all groups involved are finite and the case $r = 1$ of (*b*) then follows by counting, using Theorem 2.2 and a method

of Ogg [12]. The passage to infinite S is not difficult. As a by-product of the proof one finds that the group G has strict cohomological dimension 2 for all primes l such that $lk_S = k_S$, except of course if $l=2$ and k is not totally imaginary.

Let A be an abelian scheme over k_S and let m be a natural number such that $mk_S = k_S$. For $X = k_S$ or $X = k_P$, we put:

$$H^r(X, A; m) = \varprojlim_n [\mathrm{Coker}\,(H^r(X, A) \xrightarrow{m^n} H^r(X, A))] \quad \text{for} \quad r \leqslant 0,$$

and $$H^r(X, A; m) = \varinjlim_n [\mathrm{Ker}\,(H^r(X, A) \xrightarrow{m^n} H^r(X, A))] \quad \text{for} \quad r \geqslant 1.$$

The localization maps give homorphisms

$$H^r(k_S, A; m) \xrightarrow{\alpha_r} \prod_{P \in S} H^r(k_P, A; m),$$

where now $\prod$ denotes the (compact) *direct product* for $r=0$, and the (discrete) *direct sum* for $r \geqslant 1$. By Theorem 2.3 we have isomorphisms

$$\prod_{P \in S} H^r(k_P, A; m) \approx \Big[\prod_{P \in S} H^{1-r}(k_P, A'; m)\Big]^*$$

and consequently by duality we have maps $\beta_r = (\alpha'_{1-r})^*$:

$$\prod_{P \in S} H^r(k_P, A; m) \xrightarrow{\beta_r} H^{1-r}(k_S, A'; m)^*.$$

Let $\mathrm{Ker}^r(k_S, A; m)$ denote the kernel of α_r. For $r \geqslant 1$ and fixed m and A, this group is independent of S, by Theorem 2.5. Hence, $\mathrm{Ker}^1(k_S, A, m)$ is the m-primary component of the group of everywhere locally trivial principal homogeneous spaces for A over k. As is well known (and follows for example from [10], Theorem 5) this group is an extension of a finite group by a divisible group of "finite rank". There is canonical pairing

$$(**) \qquad \mathrm{Ker}^1(k_S, A; m) \times \mathrm{Ker}^1(k_S, A'; m) \to \mathbf{Q}/\mathbf{Z}$$

which can be defined either by a method using finite modules of m-primary division points and quite analogous to the definition of the pairing (*) above, or else by generalizing the method used by Cassels [3] in case $\dim A = 1$, the generalization involving the "reciprocity law" of Lang [9].

THEOREM 3.2. *The pairing* (**) *annihilates only the divisible part of* $\mathrm{Ker}^1\,(k, A; m)$, *nothing more.*

In case $\dim A = 1$ this theorem is due to Cassels [3], and his methods suffice for the case of general A, once one has Theorem 3.1 at one's disposal. Cassels' proof of skew symmetry in dimension 1 gives in the general case:

THEOREM 3.3. *If E is a divisor on A rational over k, and $\varphi_E : A \to A'$ the corresponding homomorphism, defined by $\varphi_E(a) = Cl(E_a - E)$, then for any $\alpha \in \mathrm{Ker}^1(k_S, A; m)$ the elements α and $\varphi_E(\alpha)$ annihilate each other in the pairing* (**).

There is also a canonical pairing

$$(***) \qquad \mathrm{Ker}^0(k_S, A; m) \times \mathrm{Ker}^2(k_S, A', m) \to \mathbf{Q}/\mathbf{Z}$$

and we have

THEOREM 3.4. *The map α_2 is surjective and its kernel is the divisible part of $H^2(k_S, A; m)$. Moreover, the following statements are equivalent:*

(i) $\mathrm{Ker}^1(k_S, A; m)$ *is finite (i.e. its divisible part is* 0).

(ii) *Im* $\alpha_0 = \mathrm{Ker}\ \beta_0$, *and the pairing* (***) *gives a perfect duality between the compact group* Ker^0 *and the discrete group* Ker^2.

Thus *if* these equivalent conditions (i) and (ii) are satisfied, we have an exact sequence

$$\begin{array}{ccccccc}
 & & 0 \to \prod_{P \text{ real}} H^2(k_P, A'; m)^* & \xrightarrow{\alpha_2^*} & H^2(k_S, A'; m)^* & \to & H^0(k_S, A; m) \\
 & & & & & & \downarrow \alpha_0 \\
\prod_{P \in S} H^1(k_P, A, m) & \xleftarrow{\alpha_1} & H^1(k_S, A; m) & \leftarrow & H^1(k_S, A'; m)^* & \xleftarrow{\beta_0} & \prod_{P \in S} H^0(k_P, A; m) \\
\beta_1 \downarrow & & & & & & \\
H^0(k_S, A'; m)^* & \to & H^2(k_S, A; m) & \xrightarrow{\alpha_2} & \prod_{P \text{ real}} H^2(k_P, A; m) & \to & 0
\end{array}$$

quite analogous to (3.1), but with the appropriate shift of dimensions by 1.

4. Conjectures

In view of Theorem 3.4, one would have to be more pessimistic than I not to make the following

CONJECTURE 4.1. $\mathrm{Ker}^1(k_S, A; m)$ *is finite.*

There is some numerical evidence for this. For example, Selmer [13] has shown $\mathrm{Ker}^1(\mathbf{Q}, A, 3)$ is finite for all but a few elliptic curves A of the form $x^3 + y^3 = cz^3$ with $0 \leqslant c \leqslant 500$. A proof of 4.1 which yielded an a priori estimate for the order of $\mathrm{Ker}^1(k_S, A; m)$ would yield an effective procedure for computing the rank of, and finding generators for, the group of rational points on an abelian variety. In general, conjecture 4.1, is in the nature of an existence theorem for rational points of infinite order on abelian varieties.

Another conjecture in the same direction is that of Birch and Swynnerton-Dyer, discussed by Cassels in his talk at this Congress, to the effect that the rank of the group of rational points on an abelian variety A of dimension 1 is determined by the order of the pole of the zeta-function of A at $s = 1$. In case (F), i.e., if k is a function field over a finite field k_0 with q elements, the two conjectures are conjecturally equivalent. Namely, let Y be the complete non-singular model of k/k_0, and X the unique complete non-singular model of $k(A)/k_0$ which is minimal with respect to the morphism $X \to Y$. Let $\overline{X} = X \times \bar{k}_0$ be the variety obtained by extending the finite ground field to its algebraic closure, and let $\varphi: \overline{X} \to \overline{X}$ be the Frobenius morphism of $\overline{X}$ relative to k_0. Combining the result of Ogg [12] and Shafaryevitch [14] with recent results of M. Artin [1] on the Grothendieck cohomology of algebraic surfaces, one sees that conjecture 4.1 is equivalent in case (F) to

CONJECTURE 4.2. *The operator $\varphi - q$ annihilates exactly that part of $H^2(\bar{X}, \mathbf{Z}_m)$ which is "algebraic" and rational over k_0, and no more.*

Clearly, 4.2 makes sense for any complete non-singular surface X over a finite field, not only for a pencil of elliptic curves. So generalized, conjecture 4.2 is equivalent, modulo Weil's well-known conjectures, to the following function theoretic analog of the conjecture of Birch and Swynnerton-Dyer.

CONJECTURE 4.3. *Let X be a complete non-singular algebraic surface defined over a finite field k_0. Then the order of the pole of the zeta-function of X at the point $s=1$ is equal to the number of algebraically independent divisors on X rational over k_0, i.e., to the k_0-picard number of X.*

Mumford has called my attention to the following interpretation of 4.3 in the special case when X is the product of two curves, one of which is elliptic.

CONJECTURE 4.3′. *Let E and E' be two complete non-singular curves defined over a finite field k_0, and suppose E is of genus 1. Then there exists a non-constant rational map $E' \to E$ defined over k_0 if and only if the zeta-function of E divides that of E'.*

In particular, if E and E' have genus 1, then they are isogenous over k_0 if and only if they have the same zeta-function. This beautiful statement has been proved by Birch and Swynnerton-Dyer and (independently) by Mumford, using results of Deuring [4] on the lifting to characteristic 0 of the Frobenius automorphism.

REFERENCES

[1]. ARTIN, M., *Grothendieck topologies.* Mimeographed notes, Harvard, 1962.

[2]. CARTAN-EILENBERG, *Homological Algebra.* Princeton Univ. Press, 1956.

[3]. CASSELS, J. W. S., Arithmetic on curves of genus 1 (IV). Proof of the Hauptvermutung, *J. reine angew. Math.*, 211 (1962), 95–112.

[4]. DEURING, M., Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abh. Math. Sem. Hamburg*, 14 (1941), 197–272.

[5]. DOUADY, A., Cohomologie des groupes compacts totalement discontinus. *Séminaire Bourbaki*, 12, 1959–60, exposé 189.

[6]. GREENBERG, M., Schemata over local rings, *Ann. Math.*, 73, 1961, 624–648.

[7]. IWASAWA, K., On Galois groups of local fields. *Trans. Amer. Math. Soc.*, 80 (1955), 448–469.

[8]. LANG, S., Algebraic groups over finite fields. *Amer. J. Math.*, 78 (1956), 555–563.

[9]. LANG, S., *Abelian Varieties.* Interscience Publishers, New York, 1959.

[10]. LANG, S. & TATE, J., Principal homogeneous spaces over abelian varieties, *Amer. J. Math.*, 80 (1958), 659–684.

[11]. NAKAYAMA, T., Cohomology of class field theory and tensor products of modules I. *Ann. Math.*, 65 (1957), 255–267.

[12]. OGG, A., Cohomology of Abelian varieties over function fields, *Ann. Math.*, 76 (1962), 185–212.

[13]. SELMER, E. S., The diophantine equation $ax^3 + by^3 + cz^3 = 0$. *Acta Math.*, 85 (1951), 203–362.

[14]. SHAFARYEVITCH, I.R., Главные однородные пространства, определенные над полем функций. *Труды Математиуеского института имени В. А. Стеклова*, Том LXIV.

[15]. SHATZ, S., *Cohomology of Artinian group schemes over local fields.* Thesis, Harvard, June, 1962.

[16]. TATE, J., W. C.-groups over P-adic fields, *Séminaire Bourbaki*, 1957–58, exposé 156.

[17]. WITT, E., Zerlegung reeller algebraischer Funktionen in Quadrate, Schiefkörper über reellen Funktionenkörper, *J. reine angew. Math.*, 171 (1934), 4–11.

TWO RESULTS ABOUT FINITE GROUPS

By JOHN G. THOMPSON

I would like to discuss two results about finite groups:

(0) All finite groups of odd order are solvable.

(F) A finite group is nilpotent if it admits a fixed point free automorphism of prime order.

Walter Feit and I proved (0) after a prolonged joint effort [5]. A critical special case of (0) was proved by Walter Feit, Marshall Hall, Jr. and me [4]. A slightly stronger result than (F) is in the literature [16].

It was Suzuki who took the first big step toward the proof of (0), for he showed that the group G is solvable if it is of odd order and if in addition the centralizer of each non-identity element of G is abelian [15]. The proof of this apparently very special case removed one of the major stumbling-blocks in the proof of (0). Suzuki's proof contains an application of exceptional characters. Exceptional characters were discovered by Brauer and Suzuki and their simplicity and power have been of great help in recent work in finite groups. Suzuki's proof also presents a structure for the proof of (0). This structure is very easy to describe; it has two parts:

(1) Determination of the maximal subgroups.

(2) Exploitation of Frobenius reciprocity.

The special case of (0) alluded to consists simply in replacing "abelian" by "nilpotent" in Suzuki's result. A skirmish with p-groups and a modification of one character argument mark the only divergence from Suzuki's proof.

The proof of (0) involves a large number of technical problems and a great deal of case analysis. Proving (0) by induction on the order of G, it can be assumed that G is simple and that every proper subgroup of G is solvable. In the determination of the maximal subgroups of G, a much more serious contact with solvable groups is required than is customary. An initial analysis, to which I will return, shows that the maximal subgroups of G are of two types. Type 1 consists of those subgroups M which are split extensions of a nilpotent normal subgroup H by a group E with the property that E contains a subgroup E_0 of the same exponent as E such that non-identity elements of E_0 induce fixed point free automorphisms of H. Furthermore, the set of elements of E which have non-identity fixed points on H lies in a normal abelian subgroup of E. Type 2 consists of those subgroups M which are extensions of a normal Hall subgroup H by a non-identity subgroup E such that any two non-identity elements of E have the same cyclic fixed point set on H. Furthermore, H is the derived group of M. These conditions imply that the derived group of H is nilpotent, and if H is not nilpotent, then E has prime order.

If M is of type 1, H_0 denotes the set of elements of M commuting with some non-identity element of H. If M is of type 2 and H is nilpotent, $H = H_0$, while if H is not nilpotent, H_0 is a suitable subset of H. The set H_0 has many properties. Two elements of H_0 are G-conjugate only if they are M-conjugate. If x is a non-identity element of H_0 and if the centralizer of x is not contained in M, then there is a maximal subgroup N of G containing the centralizer

of x with the property that $N \cap M$ has a normal complement in N. Also, the index of $N \cap M$ in N is relatively prime to the order of every element of H_0 and to certain other relevant integers.

The proofs of the preceding statements involve the "serious contact with solvable groups" mentioned above. Assuming these results, let C be the set of class functions of M vanishing outside H_0 and let C_0 be the subset of C whose elements vanish also at the identity. There is a linear isometry t mapping C_0 into the class functions of G, and t maps generalized characters into generalized characters. The existence of t is established by exhibiting explicitly suitable class functions of subgroups of G and checking that the preceding results render the induction map manageable. The mapping t is the bridge between local and global information, since if c is in C_0, then c and its image under t agree on H_0.

It is then necessary to attempt to extend t to C. The basic idea here is Feit's [3] and with a suitable modification it is shown that t may be extended provided only that certain inequalities are satisfied. These inequalities depend only on the group M and not on its imbedding in G. If these inequalities are not satisfied, the structure of M is very limited.

The most delicate portion of the above analysis deals with a subgroup M of type 2, where H is a non-abelian p-group. It is a critical part of the proof to show that even if the relevant inequalities are violated, the desired extension of t is still available. Success is obtained only by considering a suitable maximal subgroup of G which contains E and is not conjugate to M, that is, it is necessary to study the imbedding of M in G.

Once the problem of extending t has been exhaustively examined, several striking results emerge. If M is of type 1, then $E = E_0$, that is, M is a Frobenius group. If M is of type 2, then H is not nilpotent, so that E is of prime order. In conjunction with the special case of (0) already mentioned, it follows that G contains two non conjugate maximal subgroups S and T, each of type 2, whose intersection is a self-normalizing cyclic subgroup of order pq, p and q being distinct primes. Furthermore, both S' and S/S'' are Frobenius groups.

In this bizarre situation, relentless pursuit of inequalities involving sets of exceptional characters shows that both T' and T/T'' are Frobenius groups, and the structure of S and T is determined explicitly. S is isomorphic to the group of mappings $m_{a,b,s}: x \to ax^s + b$, where a, x, b are in $F(p^q)$, the field of p^q elements, $N(a) = a^{1+p+\cdots+p^{q-1}} = 1$, and where s ranges over the automorphisms of $F(p^q)$. T is obtained from S by interchanging p and q.

The symmetry between S and T is now destroyed by taking $p > q$. In this situation, T contains a cyclic subgroup V of order $(q^p - 1)/(q - 1)$ which is the centralizer of each of its non-identity elements, and V is of index pq in its normalizer. On the other hand, S contains a cyclic subgroup U of order $(p^q - 1)/(p - 1)$. Furthermore, either U is the centralizer of each of its non-identity elements and U is of index pq in its normalizer, or U is conjugate to a subgroup of V. The first possibility can be handled in a fairly straightforward fashion, since the exceptional characters for U are distinct from those for V. In the second case, however, the exceptional characters for U and V coalesce, and the character theory ceases to produce further reductions. Of course, the information that $(p^q - 1)/(p - 1)$ divides $(q^p - 1)/(q - 1)$ is now available since U is conjugate to a subgroup of V, but this information is difficult to exploit.

The final analysis involves the study of generators and relations which capture the intertwined nature of S and T and leads to a question about $F(p^q)$. Namely, if f is the linear fractional transformation sending x to $(2-x)^{-1}$, does there exist an element a of $F(p^q)$ different from 1 such that $N(a^s)=1$ for all powers s of f? This question is considerably easier than the previous divisibility problem, and the negative answer retranslates to the relations to give a contradiction and completes the proof of (0).

The problems involving solvable groups which arise in the proof of (0) are diverse. The arguments are based on P. Hall's theory of solvable groups [6–13]. The very exploitable Theorem B of Hall and Higman [14] also plays an important role. N. Blackburn has classified all p-groups which contain no normal elementary abelian subgroup of order p^3, p being an odd prime [1, 2]. Such p-groups contain no elementary abelian subgroup of order p^3. It is an easy consequence of this result that if P is such a p-group and P is a S_p-subgroup of the p-solvable group H, and H is of odd order, then H has p-length one, and elements of P are H-conjugate only if they are conjugate in the normalizer of P. Thus, for many purposes, we can consider the imbedding in G of such S_p-subgroups as known.

Let s denote the set of primes p such that G contains an elementary abelian subgroup of order p^3. If p is in s and P is a S_p-subgroup of G let A be a maximal normal abelian subgroup of P which is not generated by two elements. The first major reduction in the proof of (0) is that for each prime q different from p, the centralizer of A permutes transitively by conjugation those q-subgroups of G which are maximal with respect to the property of being normalized by A. Under the hypothesis that A normalizes some non-identity q-subgroup of G, the preceding result has a powerful consequence. Namely, if N is the normalizer of the center of the weak closure of A in P, then N has no normal subgroup of index p, and if H is any subgroup of P which contains A, then the normalizer of H is contained in N.

If p_1 and p_2 are in s, write $p_1 * p_2$ provided G contains subgroups E_1, E_2 which are elementary abelian of orders p_1^3, p_2^3 respectively, such that E_1 and E_2 generate a proper subgroup of G. The next reduction states that if $p_1 * p_2$ and if P_i is a S_{p_i}-subgroup of G, and if P_1 centralizes every p_2-subgroup of G which it normalizes and P_2 centralizes every p_1-subgroup of G which it normalizes, then G contains a nilpotent S_{p_1, p_2}-subgroup.

A variety of further reductions leads to a proof that $*$ is an equivalence relation on s and that if e is an equivalence class, then G contains a S_e-subgroup H, while $H \cap H^x$ is of square free order for every x in G which does not normalize H. The method of proof shows that H is a proper subgroup of G. The transition from this result to the more precise structure and imbedding in G of the maximal subgroups of G is somewhat technical but straightforward.

It is now possible to indicate a major difficulty in the study of the subgroups of G and at the same time to indicate the connection between this difficulty and the proof of (F). In proving (F), it was convenient to prove a stronger result. The nature of the induction hypothesis for this stronger result permits the problem to be reduced to a detailed examination of a suitable solvable subgroup K of the group in question. The group K is of order $p^a q^b$ and has the property that if H is the maximal normal p-subgroup of K and Q is the maximal normal q-subgroup of K, then H maps onto the maximal normal p-subgroup of K/Q. Furthermore, K contains a S_p-sub-

group of the normalizer of H. This is exactly the situation encountered in the proof of (0). In such a situation, it can be shown that if P is a S_p-subgroup of K contained in the S_p-subgroup P^* of G, then H contains every normal abelian subgroup of P^*. Let A be any maximal normal abelian subgroup of P^* and let C be any conjugate of A contained in K. It is important to conclude that C is contained in H, that is, Wielandt's notion of weak closure plays a dominant role here [17]. To obtain this containment, Theorem B of Hall-Higman [14] is applied. If A is not elementary, Theorem B suffices for (F). However, if A is elementary, the argument breaks down in (F), while in (0) it is only possible to assert that K is generated by $K \cap N_1$ and $K \cap N_2$, where N_1 and N_2 are subgroups of G whose definition is independent of K. Even this weaker result is quite powerful and in the proof of (0) the systematic exploitation of this fact leads eventually to the containment $C \subseteq H$. The argument succeeds because there are many subgroups which can play the role of A. It is not so important that A be a maximal normal abelian subgroup of P^*. The relevant point is that A be a normal subgroup of P^* which contains all elements of P^* which centralize A, and the class of nilpotency of A must be small enough to guarantee that $A \subseteq H$. In fact, even these conditions may be relaxed somewhat. This latitude in the choice of A plays a critical role in the proof of both (0) and (F).

In the proof of (0), the hypothesis that G is of odd order is used in a variety of ways. Theorem B is sharper for groups of odd order than for groups of even order. Various results about p-groups fail to hold for $p=2$. Non-principal irreducible characters of groups of odd order are non-real. The analysis of characters leads to inequalities in which the parameter is a prime divisor of the group order, and the inequalities force the prime to be small. If S is a group of odd order, and if S does not contain an elementary abelian subgroup of order p^3 for any prime p, then S' is nilpotent. This easy result is used frequently. The symmetric group on four letters is a counterexample if the oddness condition is omitted.

It is too early to indicate whether the techniques used in proving (0) and (F) will be relevant in the further study of finite groups. It is certain that these techniques can be refined to give a neater proof of (0) than the present one. Perhaps in conjunction with Brauer's modular theory, new results can be obtained. Also, it is reasonable that the cohomology groups of finite groups will yield insights denied to the more "internal" analysis characterizing the proofs of (0) and (F). Finally, we can hope that involutions enjoy properties which are yet unexploited.

REFERENCES

[1]. Blackburn, N., On a special class of p-groups. *Acta Math.*, 100 (1958), 45–92.

[2]. —— Generalizations of certain elementary theorems on p-groups. *Proc. London Math. Soc.*, (3) 11 (1961), 1–22.

[3]. Feit, W., Exceptional characters. *Proc. Symp. Pure Math.*, 6 (1962), 67–70.

[4]. Feit, W., Hall, M. Jr. & Thompson, J. G., Finite groups in which the centralizer of any non-identity element is nilpotent. *Math. Z.*, 74 (1960), 1–17.

[5]. FEIT, W. & THOMPSON, J. G., *Solvability of Groups of Odd Order*. (Mimeographed notes.)

[6]. HALL, P., A note on soluble groups. *J. London Math. Soc.*, 3 (1928), 98–105.

[7]. —— A contribution to the theory of groups of prime power order. *Proc. London Math. Soc.*, (2) 36 (1933), 29–95.

[8]. —— A characteristic property of soluble groups. *J. London Math. Soc.*, 12 (1937), 198–200.

[9]. —— On the Sylow systems of a soluble group. *Proc. London Math. Soc.*, (2) 43 (1937), 316–323.

[10]. —— On the system normalizers of a soluble group. *Proc. London Math. Soc.*, (2) 43 (1937), 507–528.

[11]. —— Theorems like Sylows. *Proc. London Math. Soc.*, (3), 6 (1956), pp. 286–304.

[12]. —— Some sufficient conditions for a group to be nilpotent. *Illinois J. Math.*, 2 (1958), 787–801.

[13]. —— Lecture notes.

[14]. HALL, P. & HIGMAN, G., The p-length of a p-soluble group, and reduction theorems for Burnside's problem. *Proc. London Math. Soc.*, (3) 7 (1956), 1–42.

[15]. SUZUKI, M., The nonexistence of a certain type of simple groups of odd order. *Proc. Amer. Math. Soc.*, 8 (1957), pp. 686–695.

[16]. THOMPSON, J. G., Normal p-complements for finite groups. *Math. Z.*, 72 (1960), 332–354.

[17]. WIELANDT, H., p-Sylowgruppen und p-Faktorgruppen. *J. reine angew. Math.*, 182 (1940), 180–193.

UNIQUE CONTINUATION AND LOWER BOUNDS FOR SOLUTIONS OF ABSTRACT DIFFERENTIAL EQUATIONS

By SHMUEL AGMON

1. Introduction

In recent years the uniqueness problem for the Cauchy problem for partial differential equations was considered by many authors. In particular A. P. Calderón [3] proved a very general uniqueness theorem. This has been simplified and extended by B. Malgrange [13] and L. Hörmander [6]. Prior to that various authors considered the unique continuation problem for second order elliptic equations. The most general result here was obtained by N. Aronszajn [2] and H. O. Cordes [5].

The unique continuation problem for abstract differential equations was first considered by P. D. Lax [10] who applied his results to second order elliptic equations having the Laplacian as a leading part. In the following we shall describe some recent results obtained for the abstract unique continuation problem. Our discussion will be based on a joint work of L. Nirenberg and the author [1].

We consider differential equations of the form:

$$Lu = \frac{1}{i}\frac{du}{dt} - Au = 0, \tag{1.1}$$

as well as perturbed equations (in an inequality form):

$$\|Lu\| \leqslant \Phi(t)\|u\|, \tag{1.2}$$

where $u(t)$ is a function with values in some Banach space and A is a closed, densely defined, unbounded linear operator. For simplicity it will be assumed that $u(t)$ (with values in domain of A) is strongly continuous and has a piecewise strongly continuous derivative. For solutions of (1.1) or (1.2) we shall be interested in the following uniqueness results: (i) uniqueness for the finite Cauchy problem, i.e. when $u(T)=0$ implies $u \equiv 0$ for $t<T$ (here the backward Cauchy problem is usually not a well posed problem); (ii) uniqueness at infinity. That is, when $u(t) \to 0$ with a certain rapidity as $t \to \infty$ implies that $u \equiv 0$. In some cases it will be possible to derive lower bounds for any solution of (1.1) or (1.2). This should be considered as the strongest form of a unique continuation result.

2. The non-perturbed equation

For solutions of (1.1) it is easy to obtain general conditions which ensure uniqueness. Thus for the finite Cauchy problem one has the following result (Lyubič [12] and [1]).

THEOREM 2.1. *Let $u(t)$ be a solution of* (1.1) *for* $0 \leqslant t \leqslant T$ *with* $u(T)=0$. *Assume that there is a simple Jordan curve* Γ *going to* ∞ *lying in a closed angle in the open upper half* λ*-plane on which the resolvent* $R(\lambda; A)=(\lambda-A)^{-1}$ *exists and satisfies:*

$$\|R(\lambda; A)\| = O(e^{\alpha \operatorname{Im} \lambda})$$

for some constant $\alpha \geqslant 0$. *If* $T>\alpha$ *then* $u(t)=0$ *for* $t \geqslant \alpha$.

The proof of this result follows readily taking Fourier transform and using the Phragmén Lindelöf principle.

Suppose that the operator iA is of the form:

$$iA = e^{-i\alpha}\tilde{A}, \tag{2.1}$$

where $\tilde{A}$ is the infinitesimal generator of a semi-group of bounded operators $T(t)$ and α is a real number, $0<|\alpha|<\pi$. In this case one has the following

THEOREM 2.2. *If u is a solution of* (1.1) *on* $t \geqslant 0$ *such that*

$$\|u(t)\| = O(\exp(-\varepsilon t^{\frac{\pi}{|\alpha|}})) \quad \text{as } t \to \infty,$$

for some $\varepsilon>0$, *then* $u \equiv 0$.

The proof of Theorem 2.2. is almost immediate if one observes that under the conditions on A any solution u of (1.1) in an interval (a,b) admits a continuous extension into the semi-infinite strip in the complex t-plane consisting of all points $t=s+re^{i\alpha}$ with $a<s<b$, $0 \leqslant r<\infty$, the extension being *analytic* for $a<s<b$, $r>0$. Explicitly the analytic extension is given by the formula:

$$u(t) = T(r)u(s). \tag{2.2}$$

Suppose that in (2.1) $\tilde{A}$ is actually the infinitesimal generator of a *group* $T(t)$ $(-\infty<t<\infty)$. In this case a solution of (1.1) in (a,b) is analytic in the interval and admits an analytic extension into the infinite strip: $t=s+re^{i\alpha}$, $a<s<b$, $-\infty<r<\infty$ which is given by (2.2). In the special case that $T(t)$ is uniformly bounded, say $\|T(t)\| \leqslant M$, application of Hadamard's three lines theorem in the strip shows that the following convexity result holds for any solution of (1.1) in $[a,b]$:

$$\|u(t)\| \leqslant M \|u(a)\|^{(b-t)/(b-a)} \|u(b)\|^{(t-a)/(b-a)} \quad (a<t<b).$$

From this it follows readily that any solution $u \not\equiv 0$ on $t \geqslant 0$ satisfies a lower bound of the form:

$$\|u(t)\| \geqslant ce^{-\mu t} \quad (c \text{ and } \mu>0).$$

If $T(t)$ is not uniformly bounded a somewhat more involved convexity relation shows that any solution on $t \geqslant 0$ satisfies a lower bound of the form:

$$\|u(t)\| \geqslant ct^{-\mu t}.$$

Another lower bound is obtained when one takes an intermediate situation between the two considered above. Namely, when the assumption that $\tilde{A}$ is the infinitesimal generator of a group is replaced by the assumption that

both A and $\gamma\tilde{A}$ for some non-real γ are the infinitesimal generators of semi-groups. This situation appears in the case of analytic semi-groups and was considered by Krein and Prozorovskaya [8]. Finally we note that lower bounds for solutions of (1.1) in a general case when A is not the infinitesimal generator of a semi-group were obtained in [1].

3. The finite Cauchy problem for the perturbed equation

We consider now the unique continuation problem for the perturbed equation:

$$\|Lu\| = \left\| \frac{1}{i}\frac{du}{dt} - Au \right\| \leqslant \Phi(t)\,\|u\|. \tag{3.1}$$

In this section we assume that the Banach space is a Hilbert space. The following theorem was proved by P. D. Lax [10].

THEOREM 3.1. *Let u be a solution of* (3.1). *Assume that there is a sequence of lines in the complex λ-plane parallel to the real axis,* $\operatorname{Im}\lambda = a_n$, *with* $a_n \to \infty$, *on which the resolvent* $R(\lambda; A) = (\lambda - A)^{-1}$ *exists and is bounded by a constant* M. *If* $\Phi(t) \leqslant c$ *for some* $c < M^{-1}$ *then* $\|u(t)\| = O(e^{-at})$ as $t \to \infty$ *for every positive a implies that* $u \equiv 0$.

It was shown by Lax that the condition $c < M^{-1}$ cannot be dropped. If, however, one is interested in uniqueness for the finite Cauchy problem the conditions on the resolvent could be relaxed. The following result was established in [1].

THEOREM 3.2. *Let u be a solution of* (3.1) *in an interval* $0 \leqslant t \leqslant T$ *with* $u(T) = 0$. *Let F be a sequence of lines* $\operatorname{Im}\lambda = a_n$ *in the λ-plane with* $a_n \to \infty$. *Assume that there exist positive numbers s, M and an integer $j > 0$ such that on each line of F the resolvent $R(\lambda; A)$ is bounded by M outside j intervals of length s (the intervals may vary from line to line.) Then there is a constant $c > 0$ depending only on j, s and M such that if* $\Phi(t) \leqslant c$ *then* $u \equiv 0$.

We mention also the following variant of the last theorem [1] where one imposes a stronger restriction on the resolvent but relaxes the assumption on Φ.

THEOREM 3.3. *Let u be a solution of* (3.1) *in* $[0,T]$ *with* $u(T) = 0$. *Suppose that on a sequence of lines* $\operatorname{Im}\lambda = a_n$, $a_n \to \infty$, *the following inequality holds:*

$$\|R(\lambda; A)\| \leqslant \frac{M}{|\operatorname{Re}(\lambda - \lambda_n)|^{\sigma}} \quad \text{for } |\operatorname{Re}(\lambda - \lambda_n)| \geqslant s,$$

for some point λ_n with $\operatorname{Im}\lambda_n = a_n$, *and where σ, s are positive constants,* $\sigma \leqslant 1$. *Then if Φ belongs to L_p on the interval with* $2 \leqslant p < \infty$ *and* $p > \sigma^{-1}$, *then* $u \equiv 0$.

4. Convexity results and lower bounds

We turn to the more special unique continuation results which involve lower bounds for solutions of the perturbed equation (3.1). P. Cohen and M. Lees [4] have shown that if iA is a symmetric operator in a Hilbert space

and if $\Phi(t) \in L_p(0,\infty)$, $1 \leqslant p \leqslant 2$, then any solution $u \not\equiv 0$ on $t \geqslant 0$ satisfies a lower bound of the form:

$$\|u(t)\| \geqslant ce^{-\mu t}, \tag{4.1}$$

c and μ some positive constants depending on the solution.

In case A is a non real multiple of a self-adjoint operator (4.1) and various extensions were derived in [1] from the following convexity result.

THEOREM 4.1. *Suppose that A is a non-real multiple of a self-adjoint operator. Let u be a solution of* (3.1) *in an interval* $[a,b]$ *over which*

$$\int_a^b \Phi(t)\,dt \leqslant \frac{1}{2\sqrt{2}}.$$

Then u satisfies the following convexity inequality

$$\|u(t)\| \leqslant 2\sqrt{2}\,\|u(a)\|^{(b-t)/(b-a)}\,\|u(b)\|^{(t-a)/(b-a)} \quad (a \leqslant t \leqslant b).$$

Convexity results hold not only for self-adjoint operators and not only in Hilbert spaces. The following theorem is an example of such a result in a more general situation.

THEOREM 4.2. *Suppose that the operator iA (in a Banach space) is a non-real multiple of an infinitesimal generator of group $T(t)$. Let the function Φ in* (3.1) *be a constant. Then there exist positive numbers d and C (depending only on A and Φ) such that for every solution of* (3.1) *in an interval $[a,b]$ of length $\leqslant d$ the following inequality holds:*

$$\|u(t)\| \leqslant C\,\|u(a)\|^{(b-t)/(b-a)}\,\|u(b)\|^{(t-a)/(b-a)} \quad (a < t < b). \tag{4.2}$$

We note that the last theorem admits a refinement if the group of operators satisfies an estimate of the form

$$\|T(t)\| \leqslant \text{const}\,(1+|t|^l) \quad \text{for} \quad -\infty < t < \infty \tag{4.3}$$

and some $l \geqslant 0$.

From the convexity results it is possible to derive lower bounds for solutions of (3.1) on the positive axis. The following two theorems are special cases of more general results [1].

THEOREM 4.3. *Suppose that the conditions of Theorem* 4.2. *hold and let $u \not\equiv 0$ be a solution of* (3.1) *for $t \geqslant 0$. Then*

$$\|u(t)\| \geqslant ce^{-\mu t^2} \quad \text{for} \quad t \geqslant 0, \tag{4.4}$$

c, μ some positive constants.

THEOREM 4.4. *Suppose that A satisfies the conditions of Theorem* 4.2 *and that in addition* (4.3) *holds. Let $u \not\equiv 0$ be a solution of* (3.1) *for $t \geqslant 0$, and suppose that*

$$\Phi(t) \leqslant \text{const}\,(1+t)^{-k} \quad \text{for} \quad t \geqslant 0$$

with $k > l+1$. Then:

$$\|u(t)\| \geqslant ce^{-\mu t} \quad (c, \mu > 0).$$

5. On applications

The abstract approach was used by P. D. Lax [10] to prove the unique continuation theorem for second order elliptic equations possessing the Laplacian as their principal part.

The results described here and in particular those of § 4 could be used to derive lower bounds (via a convexity argument) for solutions of parabolic and elliptic problems in semi-infinite cylindrical domains. In this connection we note that convexity results for (perturbed) parabolic problems were obtained by Krein [7], and for elliptic problems by Lavrentev [9]. The unique continuation problem at infinity for parabolic problems was considered by Protter [14] and Lees [11].

REFERENCES

[1]. Agmon, S. & Nirenberg, L., Properties of solutions of ordinary differential equations in Banach space. *Comm. Pure Appl. Math.* (To appear.)

[2]. Aronszajn, N., A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order. *J. Math. Pure App.*, 36 (1957), 235–249.

[3]. Calderón, A. P., Uniqueness in the Cauchy problem for partial differential equations. *Amer. J. Math.*, 80 (1958), 16–36.

[4]. Cohen, P. J. & Lees, M., Asymptotic decay of solutions of differential inequalities. *Pacific J. Math.* 11 (1961), 1235–1249.

[5]. Cordes, H. O., Über die Bestimmtheit der Lösungen ellipticher Differentialgleichungen durch Anfangsvorgaben. *Nach. Akad. Wiss. Göttingen Math.-Phys.*, 11 (1956), 239–258.

[6]. Hörmander, L., On the uniqueness of the Cauchy problem. *Math. Scand.*, 6 (1958), 213–225; part II, 7 (1959), 177–190.

[7]. Krein, S. G., On some classes of correctly posed boundary value problems. *Dokl. Akad. Nauk, S.S.S.R.*, 114 (1957), 1162–1165.

[8]. Krein, S. G. & Prozorovskaya, O. I., Analytic semi-groups and incorrect problems for evolutionary equations. *Dokl. Akad. Nauk, S.S.S.R.*, 133 (1960), 277–280.

[9]. Lavrentev, M. M., On Cauchy's problem for linear elliptic equations of the second order. *Dokl. Akad. Nauk, S.S.S.R.*, 112 (1957), 195–197.

[10]. Lax, P. D., A stability theorem for solutions of abstract differential equations, and its application to the study of the local behaviour of solutions of elliptic equations. *Comm. Pure Appl. Math.*, 9 (1956), 747–766.

[11]. Lees, M., Properties of solutions of differential inequalities. (To appear.)

[12]. Lyubič, Ju. I., Conditions for the uniqueness of the solution to Cauchy's abstract problem. *Dokl. Acad. Nauk, S.S.S.R.*, 130 (1960), 969–972.

[13]. Malgrange, B., Unicité du problème de Cauchy d'après A. P. Calderón. *Séminaire Bourbaki*, Feb., 1959, 178.

[14]. Protter, M. H., Properties of solutions of parabolic equations and inequalities. *Can. J. Math.*, 13 (1961), 331–345.

COMPLEX PSEUDOCONCAVE SPACES AND AUTOMORPHIC FUNCTIONS

By ALDO ANDREOTTI

Compact complex spaces and holomorphically complete spaces have been widely studied in recent years. An attempt to fill the gap between these two somewhat extreme instances has been made recently [3] with the introduction of the notion of q-pseudoconvex and q-pseudoconcave spaces. This theory is at its beginning, and applications are not very numerous. We will deal here with a very special case, the case of "pseudoconcave spaces", which has shed some light on the theory of automorphic functions fo several complex variables. We will try to give a self-contained outline of the subject (cf. [2]) in the case which is of interest for that theory (for more comprehensive theorems we refer to [1]).

(a) Let D be an open set in $\mathbf{C}^n$, and let $\partial D = \bar{D} - D$ be its boundary.

We say that D is *pseudoconcave at a boundary point* $z_0 \in \partial D$ if there exists a fundamental system of open neighborhoods $\{U\}$ of z_0 in $\mathbf{C}^n$ such that $(\widehat{U \cap D})_{\mathcal{H}(U)}$ contains z_0 as an interior point.

Here $\mathcal{H}(U)$ is the algebra of holomorphic functions on U and we have denoted for any set $V \subset U$

$$\hat{V}_{\mathcal{H}(U)} = \{z \in U \mid \ |f(z)| \leqslant \sup |f(V)| \quad \text{for every} \quad f \in \mathcal{H}(U)\}.$$

A criterion for deciding the psuedoconcavity at a boundary point z_0 is the following:

The domain D is pseudoconcave at $z_0 \in \partial D$ if there exists a two-complex-dimensional plane $E(t_1, t_2)$ through z_0, a neighborhood V of z_0 in E and a C^∞ function p on E such that

$$p(z_0) = 0, \ V \cap D = \{t \in V \mid p(t) < 0\},$$

$$L(p) = \sum \frac{\partial^2 p}{\partial t^\alpha \partial \bar{t}^\beta} dt^\alpha \, d\bar{t}^\beta \quad \textit{is positive definite at } z_0.$$

b) Let X be a complex connected manifold (with countable topology) of complex dimension n, and let Γ be a group of complex analytic automorphisms of X.

DEFINITION. *We say that Γ is pseudoconcave if there exists a non-empty relatively-compact open subset $D \Subset X$ such that for every point $z_0 \in \partial D$ there exists in the orbit Γz_0 either an interior point of D or a boundary point of D where D is pseudoconcave.*

If Γ is properly discontinuous (i.e. for every compact $K \subset X$ the set $\{\gamma \in \Gamma \mid \gamma K \cap K \neq \phi\}$ is finite) and X/Γ is compact, then Γ is pseudoconcave.

If Γ_1 and Γ_2 are two groups of automorphisms of X and if $\Gamma_1 \cap \Gamma_2$ is of finite index in Γ_1 and Γ_2, we say that Γ_1 and Γ_2 are commensurable. If Γ_1 and Γ_2 are commensurable and if Γ_1 is pseudoconcave, Γ_2 is also pseudoconcave.

If Γ is pseudoconcave, then any Γ-invariant holomorphic function on X is constant.

(*c*) An automorphic line bundle F over X is a holomorphic line bundle on which Γ acts by automorphisms compatible with the projection map on X.

If X is a contractile Stein manifold, the datum of an automorphic line bundle is equivalent to the datum of a factor of automorphy.

Given an automorphic line bundle F over X with respect to Γ, we can speak of the space of Γ-invariant sections of F. The main theorem is the following:

THEOREM. *If Γ is pseudoconcave, then the space of Γ-invariant sections of F is of finite dimension.*

Let F^h be the hth tensor power of the line bundle F; it is also an automorphic line bundle. From the proof of the above theorem one gets for the dimension of the space $H^\circ(X, F^h)^\Gamma$ of Γ-invariant sections of F^h an estimate of the following type:

$$\dim_{\mathbf{C}} H^\circ(X, F^h)^\Gamma \leqslant \text{polynomial in } h \text{ of degree} \leqslant n.$$

(*d*) Assume now that Γ is properly discontinuous. Consider the ring

$$\mathcal{A}^\Gamma(F) = \bigcup_h H^\circ(X, F^h)^\Gamma.$$

This is a graded integral domain. Let $Q^\Gamma(F)$ be its field of quotients and let $\mathcal{K}^\Gamma(X)$ be the field of Γ-invariant meromorphic functions on X; then obviously

$$Q^\Gamma(F) \subset \mathcal{K}^\Gamma(X).$$

From the above considerations one deduces the following

PROPOSITION. *The transcendence degree of $\mathcal{K}^\Gamma(X)$ is $\leqslant \dim_{\mathbf{C}} X$; for every automorphic line bundle F the field $Q^\Gamma(F)$ is algebraically closed in $\mathcal{K}^\Gamma(X)$; the ring $\mathcal{A}^\Gamma(F)$ is integrally closed in $Q^\Gamma(F)$.*

In particular if trans. deg. $Q^\Gamma(F) = \dim_{\mathbf{C}} X$, then every Γ-invariant meromorphic function on X can be written as a quotient of two sections of a tensor power of the line bundle F; the field $\mathcal{K}^\Gamma(X)$ is then isomorphic to the field of rational functions on a projective algebraic variety of complex dimension n.

Consider now the space X/Γ; this is a complex space in a natural way [5]. *If X/Γ can be imbedded as a locally closed subset of a projective space $P_N(\mathbf{C})$, then X/Γ, so imbedded, is a part of an algebraic irreducible variety of $P_N(\mathbf{C})$ of the same dimension.*

(*e*) We conclude with the example of the modular groups of degree $n > 1$ (cf. [7]). Here the space X is the generalized upper half plane

$$X = \{Z^{(n,n)} \in \mathbf{C}^{n^2} \mid {}^tZ = Z \quad \operatorname{Im} Z > 0\}.$$

This space is a contractile domain of holomorphy of $\mathbf{C}^{n(n+1)/2}$. The group is the modular group of automorphisms of X:

$$Z \to (A\,Z+B)\ (C\,Z+D)^{-1},$$

where $M=\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is a matrix with integral elements satisfying the condition

$${}^tM\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}M=\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Let $Z=X+iY$, $X=(x_{\mu\nu})$, and write Y in the Jacobi form: $Y={}^tW\Delta W$ where $\Delta=\operatorname{diag}(d_1,\ldots,d_n)$ and $W=\begin{pmatrix} 1 & & w_{ij} \\ & \ddots & \\ o & & 1 \end{pmatrix}$.

Consider the fundamental open set for Γ:

$$\Omega_u=\begin{cases} |x_{\nu\mu}|<u \\ |w_{\nu\mu}|<u \\ 1<u\,d_1<\ldots<ud_n \end{cases} \qquad (u \text{ large}).$$

The set

$$D_{uk}=\{Z\in\Omega_u \mid \det.\ Y<k\}$$

is relatively compact in X and for large k one can prove, taking $D=D_{uk}$ and using the criterion given in (a), that Γ is pseudoconcave (cf. [2]).

In particular one deduces that the field of Γ-invariant meromophic functions on X is an algebraic function field, a theorem first proved by Baily [4] using the compactification of Satake [6] of the quotient space X/Γ.

A more general class of examples is given by the theory of Riemann matrices as it has been remarked by Ramanhattan.

REFERENCES

[1]. Andreotti, A., Théorèmes de dépendence algébrique sur les espaces complexes pseudoconcaves. *Bull. Soc. Math. France.* (To appear.)

[2]. Andreotti, A. & Grauert, H., Algebraische Körper von automorphen Funktionen. *Nachr. Akad. Wiss. Göttingen,* (1961), 39–48.

[3]. —— Théorèmes de finitude pour la cohomologie des espaces complexes. *Bull. Soc. Math. France,* 90 (1962).

[4]. Baily, W. L., Satake compactification of V_n^*. *Amer. J. Math.,* 80 (1958), 348–364.

[5]. Cartan, H., Quotient d'un espace analytique par un groupe d'automorphismes. *Algebraic Geometry and Topology, A Symposium in Honor of S. Lefschetz,* 90–102. Princeton University Press, 1957.

[6]. Satake, J., On the compactification of the Siegel space. *J. Indian Math. Soc.,* 20 (1956), 259–281.

[7]. Siegel, C. L., Einführung in die Theorie der Modulfunktionen n-ten Grades. *Math. Ann.,* 116 (1939), 617–657.

ON THE MODULI OF ABELIAN VARIETIES WITH MULTIPLICATIONS FROM AN ORDER IN A TOTALLY REAL NUMBER FIELD(1)

By WALTER L. BAILY, JR.

1. Introduction

We wish to present here a theory of moduli for connected families of polarized Abelian varieties whose endomorphism ring contains a given order in a totally real algebraic number field. Our results overlap results of G. Shimura [5] and results of S. Katayama. In certain cases we are able to employ the theory of θ-functions to obtain results more precise than those of Shimura and Katayama concerning generation of the field of moduli by certain automorphic functions. The curious fact is that the results of these authors, insofar as they are not concerned with the Fourier expansions of automorphic forms, coincide with ours up to a set of measure zero, and in their work the θ-functions do not play such a dominant role. Moreover, the work of Shimura on the families of Abelian varieties whose endomorphism rings contain a maximal order of an indefinite quaternion algebra makes very little use of θ-functions. This leads us to suspect that our methods could be substantially simplified in their utilization of θ-functions. However, we should not lose sight of the essential nature of the results described here (and of the results of the above authors) as opposed to results of a more purely algebraic nature on the general problem of moduli. Namely, the results presented in this communication describe a concrete model for the variety of moduli of Abelian varieties with multiplications as the quotient of a symmetric Riemannian space by a discrete group of isometries, and tell us a set of transcendental functions which describe the field of moduli.

In presenting the results that follow, we wish to emphasize that in the preparation of these results we have had access to a prepublication copy of a manuscript of S. Katayama which has been very useful in clarifying the situation when the order in the number field is maximal and in focussing our attention on the essential algebraic difficulties which arise in the case of non-maximal orders. Moreover, the results of I. I. Pyateckii-Shapiro [4] on singular modular functions have also been very useful to us.

2. Definitions

By an analytic family of Abelian varieties we mean a triple (A,λ,B), where A and B are irreducible complex analytic spaces and λ is a proper complex analytic mapping of A onto B such that for some proper analytic subset $\mathcal{E}$ of B, and for any $b\in B-\mathcal{E}$, the "fiber" $A_b=\lambda^{-1}(b)$ is an Abelian variety of dimension n (fixed n for all $b\in B-\mathcal{E}$), and such that, putting $A^{(l)}=(\lambda^l)^{-1}$ (diagonal of B^l) and $\lambda^{(l)}=(\lambda^l)|A^{(l)}$ (for any positive integer

(1) Research for this paper was supported in part by funds from the National Science Foundation and from the Alfred P. Sloan Foundation.

l), the group law of A_b is cut out on $A_b^{(3)} = \lambda^{(3)-1}(b)$ by an analytic subset G of $A^{(3)}$ for all $b \in B - \mathcal{E}$. Let r be some ring. We say that (A, λ, B) admits the ring r as an endomorphism ring if for each $\varrho \in r$ we are given an analytic subset $\iota(\varrho)$ of $A^{(2)}$ such that for each $b \in B - \mathcal{E}$, $\iota_b(\varrho) = \iota(\varrho) \cap A_b^{(2)}$ is the graph of an endomorphism of A_b and such that

$$\iota_b : \varrho \to \iota_b(\varrho)$$

is an isomorphism of r into $\mathcal{A}(A_b)$ for all $b \in B - \mathcal{E}$.

The same definitions will also be used with the usual replacement of complex analytic objects by the corresponding objects of algebraic geometry. If k is a field, we say (A, λ, B) is defined over k if A, λ, and B are defined over k. We say $(A, \lambda, B, \iota, r)$ is defined over k if (A, λ, B) and every element of $\iota(r)$ are defined over k.

3. Reduction of the problem

Given any algebraic family of Abelian varieties in some projective space, the polarization on each fiber $A_b, b \in B - \mathcal{E}$, attached to a hyperplane section, gives rise to an involution $*$ of the endomorphism algebra of A_b. Given a projective algebraic system $(A, \lambda, B, \iota, r)$ where r is an order in a totally real algebraic number field F, we shall assume that $*$ is the identity on $\iota_b(r)$. Moreover, if 1 is the identity of r, we assume that $\iota_b(1)$ is the identity of $\mathcal{A}(A_b)$. Let $\dim A_b = n$ for each $b \in B - \mathcal{E}$, and let s be the degree of F over the rational numbers Q. Then s divides n. Let $n = sp$.

Denote by H_p the space of $p \times p$ symmetric complex matrices $Z = X + iY$ such that Y is positive definite. By a lattice in F^l we mean a free Abelian subgroup L of F^l having sl independent generators. Let $Z = (Z_1, \ldots, Z_s) \in H_p^s$, denote the $p \times p$ identity matrix by E, and let L be a lattice in F^{2p}. Denote by L_Z the lattice in C^n consisting of all the vectors

$$((Z_1 E)\xi^{(1)}, \ldots, (Z_s E)\xi^{(s)}) \quad (\xi \in L),$$

where $(Z_i E)$ denotes a $p \times 2p$ matrix written in $p \times p$ blocks and $\xi^{(i)} = \xi^{\sigma_i}$, $i = 1, \ldots, s$, denote the conjugates of ξ over Q, $\sigma_1 \ldots, \sigma_s$ denoting the distinct isomorphisms of F into the reals (ξ, etc., are viewed as column vectors). C^n/L_Z is an Abelian variety, and the skew-symmetric form defined on pairs (x, y) of s-tuples of $2p$-dimensional (real) vectors $x = (x_1, \ldots, x_s), y = (y_1, \ldots, y_s)$ defined by $S(x, y) = \sum_1^s {}^t x_i J y_i$, where $J = \left(\begin{smallmatrix} 0 & E \\ -E & 0 \end{smallmatrix}\right)$, defines a principal form for L_Z. (Observe that viewed as a skew-symmetric form on the lattice L considered as the subgroup of integral vectors of Q^{2n} the elementary divisors of S are not in general equal thus the polarization supplied by S is not in general the normal polarization.) Define an equivalence relation $\sim$ on $H_p^s \times (C^p)^s = H_p^s \times C^n$ by: $(Z, \zeta) \sim (Z', \zeta')$ if and only if $Z = Z', \zeta - \zeta' \in L_Z$. Then the quotient A of $H_p^s \times C^n$ by this equivalence relation together with $B = H_p^s$ and the projection λ of A onto B is a fiber system of Abelian varieties, admitting in a natural way any subring of the order of L as a ring of multiplications with the mapping ι defined in an obvious manner. Then if $(A', \lambda', B', \iota', r)$ is any fiber system of Abelian varieties with endomorphisms as described at the

beginning of this section, it is easy to prove that for some lattice L in F^{2p}, $(A', \lambda', B', \iota', r)$ is locally induced by $(A, \lambda, B, \iota, r)$ in some neighborhood of each point of $B' - \mathcal{E}'$. We henceforth assume that $(A, \lambda, B, \iota, r)$ is given as above.

4. Methods and results

L being a lattice as in § 3, assume that $L = L_1 \oplus L_2$, where $L_1 = \{a \in F^p \mid (a, 0) \in L\}$, $L_2 = \{b \in F^p \mid (0, b) \in L_2\}$. If r is the maximal order o of k, then r is a Dedekind ring and Katayama has proved that, with appropriate choice of coordinates in F^{2p}, we may always assume this is true. In the general case, it may or may not be possible to make this assumption, depending on whether or not a certain element τ of some group $Ext_r(M, N)$ vanishes, for certain r-modules M and N. Since the latter is properly concerned with homological algebra, we omit further discussion of it here, except to say that it appears there may be some connection between τ and the smallest field of definition for the variety of moduli (this field of definition is in any case contained in some cyclotomic field).

We now consider θ-functions suitable for imbedding $A_Z = C^n/L_Z$ in projective space. With $\sigma_1, \ldots, \sigma_s$ having the same meaning as before, and for $\zeta = (\zeta_1, \ldots, \zeta_s) \in (C^p)^s$ and $Z = (Z_1, \ldots, Z_s) \in H_p^s$, define $\zeta^{\sigma_i} = \zeta_i, Z^{\sigma_i} = Z_i$. For any square matrix M of any number of rows define $tr(M) = \sum_i M^{\sigma_i}$. Then if $g, h \in F^p, \zeta \in C^n$, and $Z \in H_p^s$, define:

$$\theta[g, h](\zeta, Z) = \sum_{\xi_1 \in L_1} \varepsilon(\tfrac{1}{2} Z[\xi_1 + g] + {}^t(\xi_1 + g)(\zeta + h)),$$

where $\varepsilon(\ ) = e^{2\pi i \cdot tr(\ )}$ and $Z[a] = {}^t a Z a$. If L_3 is any lattice in F^p, let $L_3' = \{a \in F^p \mid tr({}^t a L_3) \subset \mathrm{Z}\}$, Z denoting the rational integers; L_3' is called the complementary lattice of L_3. Without loss of generality, we may assume in the above situation that $L_2 \supset L_1'$. (We note that any two lattices in F^l are commensurable.) Then $L_2' \subset L_1$. By a θ-function of mth order with respect to L, we mean an entire function θ of ζ (for fixed Z) satisfying

$$\theta(\zeta + Z\xi_1 + \xi_2) = \varepsilon(-\tfrac{1}{2} m Z[\xi_1] - m\, {}^t\xi_1 \zeta)\, \theta(\zeta), \xi_1 \in L_1, \xi_2 \in L_2.$$

If m is a positive integer such that $mL_2 \subset L_1'$, the module of such functions is not $\{0\}$ and a basis is given by the functions $\theta[\eta_2/m, 0](m\zeta, mZ)$, $\eta_2 \in L_2$ modulo $mL_1 \subset L_2'$. The module of such θ-functions is also spanned by the products:

$$\prod_{i=1}^{m} \theta \left[\frac{\eta_{2i}}{m}, \frac{\eta_{1i}}{m}\right](\zeta, Z), \quad \eta_{2i} \in L_2', \quad \eta_{1i} \in L_1', \quad \sum \eta_{1i} = \sum \eta_{2i} = 0.$$

We denote by Γ_L the group of symplectic $2p \times 2p$ matrices M over F such that

$${}^t M^{-1} L = L$$

Then Γ_L operates on $(H_p)^s$ naturally [4] and it is easy to prove that the points of H_p^s/Γ_L are in a natural one-to-one correspondence with the isomorphism classes of such polarized Abelian varieties with multiplications.

For suitable large m, let $\mathfrak{M}$ denote the module of homogeneous forms of suitable high degree d (we assume that $2|d$) in the functions:

$$\theta\left[\frac{a}{d}, \frac{b}{d}\right](0, Z), \quad a \in L_2', \quad b \in L_1'.$$

$\mathfrak{M}$ consists of automorphic forms with respect to a group Γ_L^* contained in some suitable congruence subgroup $\Gamma_L(l) = \{\gamma \in \Gamma_L \mid {}^t\gamma^{-1}\xi - \xi \in lL$ for all $\xi \in L\}$ of Γ_L, $l > 2$. Thus no element of Γ_L^* other than the identity has fixed points in H_p^s. Put

$$B^* = H_p^s/\Gamma_L^*, \; A^* = A/\Gamma_L^*,$$

$$\lambda^* : A^* \to B^*, \quad \text{the natural map.}$$

Then $(A^* \lambda^* B^*)$ is an analytic fiber system of Abelian varieties (with $\mathcal{E}^* = \emptyset$) which may be completed to a projective algebraic system of Abelian varieties by adding analytic sets of lower dimension. The projective imbedding may be achieved by means of the θ-functions discussed above and by modular forms in the integral closure of the graded ring generated by $\mathfrak{M}$. All the projective varieties under consideration are defined over the cyclotomic field $F_d = Q(e^{\pi i/d^2})$. The Fourier coefficients of the above functions lie in this field, and if σ belongs to the Galois group of F_d/Q, and $\sigma(e^{\pi i/d^2}) = e^{s\pi i/d^2}$ with $(s, d) = 1$, define $\sigma\theta[a/d, b/d] = \theta[a/d, sb/d]$. The effect of this is the same as applying σ to each of the Fourier coefficients of $\theta[a/d, b/d]$. It is clear that σ permutes the different θ-functions under consideration and induces a linear transformation of the module $\mathfrak{M}$ and of the module of mth order θ-functions. Thus σ defines an isomorphism T_σ defined over F_d of $(A^* \lambda^* B^* \iota^* r)$ onto $(A^* \lambda^* B^* \iota^* r)^\sigma$. If $\gamma \in \Gamma_L$, we may use the general theory of the transformation of θ-functions to show that γ induces a projective transformation T_γ defined over F_d of $(A^*, \lambda^*, B^*, \iota^*, r)$ onto itself. Both T_γ and T_σ transform the fibers A_b^* into isomorphic polarized Abelian varieties with endomorphism ring containing r. From this, using methods of a previous paper of ours [1] and a criterion of A. Weil [6], we obtain the following result:

There exists an imbedding ϕ of H_p^s/Γ_L into some complex projective space CP^N by means of automorphic functions with rational Fourier coefficients such that $\phi(H_p^s/\Gamma_L)$ is a Q-open subset of an algebraic variety defined over Q and such that if $x \in H_p^s/\Gamma_L$, then $Q(\phi(x))$ is precisely the field of moduli of the Abelian variety with multiplications corresponding to x.

Using established results for the Hermitian modular functions [3], we feel that the same kind of methods should yield an analogous result in the case of the moduli of Abelian varieties with endomorphism ring containing an order in some algebraic number field of CM-type. However, we do not see at present how to extend these methods to cases such as that considered by Shimura for indefinite quaternion algebras [5] where the fundamental domain of the discontinuous group is compact. We feel it would be desirable to unify the methods in the two cases of when the fundamental domain is compact and when it is not. We also feel it would be of considerable interest to find a more complete answer in the case considered by us when the order in F is not the maximal order.

REFERENCES

[1]. BAILY, JR., W. L., On the theory of θ-functions, the moduli of Abelian varieties, and the moduli of curves. *Ann. Math.*, 75 (1962), 342–381.

[2]. —— On the Hilbert–Siegel modular space. *Amer. J. Math.*, 81 (1959), 846–874.

[3]. KLINGEN, H., Diskontinuierliche Gruppen in Symmetrischen Räumen I, II. *Math. Ann.*, 129 (1955), 345–369; 130 (1955), 137–146.

[4]. PYATECKII-SHAPIRO, I. I., Singular modular functions. *Izv. Akad. Nauk USSR Ser. Mat.*, 20 (1956), 53–98.

[5]. SHIMURA, G., On the theory of automorphic functions. *Ann. Math.*, 70 (1959), 101–144.

[6]. WEIL, A., The field of definition of a variety. *Amer. J. Math.*, 78 (1956), 509–524.

INTERPOLATIONS BY BOUNDED ANALYTIC FUNCTIONS AND THE CORONA PROBLEM(1)

By LENNART CARLESON

1. Let H^p, $1 \leqslant p \leqslant \infty$, denote the Banach space of analytic functions in $|z| < 1$ under the norm

$$\|G\|_p = \begin{cases} \lim\limits_{r \to 1} \left\{ \dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} |G(re^{i\theta})|^p \, d\theta \right\}^{1/p} & (1 \leqslant p \leqslant \infty), \\ \sup |G(re^{i\theta})| & (p = \infty). \end{cases}$$

We shall be concerned with the interpolation problem, $|w_\nu| \leqslant 1$,

$$f(a_\nu) = w_\nu \quad (\nu = 1, 2, \ldots, n, \; f \in H^\infty). \tag{1}$$

In [1], it was proved that under the condition (1)

$$\inf \|f\|_\infty = \sup_G \left| \sum \frac{G(a_\nu) w_\nu}{A'(a_\nu)} \right| \quad (\|G\|_1 = 1). \tag{2}$$

Here $A(z)$ denotes the Blaschke product

$$A(z) = \prod_1^n \frac{a_\nu - z}{1 - z\bar{a}_\nu}.$$

We denote by $a(\varepsilon)$ the set

$$a(\varepsilon) = \{z \,|\, |A(z)| \leqslant \varepsilon\},$$

and consider the particular case when

$$w_\nu = F(a_\nu)$$

for a function $F(z)$ analytic at each point of $a(\delta)$, $|F(z)| \leqslant 1$. We can then use the Cauchy formula in (2). If the curve Γ is situated in $a(\delta) - a(\varepsilon)$ we obtain the estimate

$$\inf \|f\|_\infty \leqslant \sup_G \frac{1}{2\pi\varepsilon} \int_\Gamma |G(z)| \, |dz|. \tag{3}$$

To estimate the right hand side we use the following theorem.

THEOREM 1. *Let $\mu(z)$ be a positive measure in $|z| < 1$ and denote by S any subset of $|z| < 1$ of the following form*

$$S = \{z \,|\, 1 - q \leqslant |z| < 1, \theta_0 \leqslant \arg z \leqslant \theta_0 + q\} \quad (q < 1).$$

(1) This work has been supported by a grant (G14600) from the National Science Foundation and by the U.S. Air Force, European Division under contract no. AF 61(052)–238.

Then
$$\mu(S) \leqslant Cq \tag{4}$$
for all S implies
$$\iint |G(z)|^p d\mu(z) \leqslant AC\|G\|_p^p$$
for a certain absolute constant A.

Remark. The Hilbert double series inequality and the Hardy-Littlewood maximal theorem are special cases (except for the value of the constants) of Theorem 1.

It can now be proved that there exists an absolute constant $\varkappa > 0$ such that, independently of the choice of $A(z)$, every set $a(\varepsilon^\varkappa) - a(\varepsilon)$ contains a curve Γ such that arc length along Γ satisfies (4), with $C = O(\varepsilon^{-2})$. This means that the right-hand side of (3) is uniformly bounded and that we have obtained the following theorem.

THEOREM 2. *If $F(z)$ is analytic in $a(\delta)$ and $|F| \leqslant 1$, we can solve the interpolation $f(a_\nu) = F(a_\nu)$ with*
$$\|f\|_\infty < \delta^{-K} \quad (K \textit{ absolute constant}).$$

This theorem has as simple consequence a proof of the so called Corona conjecture for H^∞. Let $\mathfrak{M}$ denote the maximal ideal space of H^∞. The part of $\mathfrak{M}$ which can be identified with the unit disc is U. Then
$$\bar{U} = \mathfrak{M}.$$
The deduction of this from Theorem 2 is due to D. J. Newman, who also conjectured Theorem 2.

It should be observed that Theorem 2 also contains the main result of [1], namely that the interpolation $f(a_\nu) = w_\nu$ is possible for arbitrary w_ν, $|w_\nu| \leqslant 1$, with $f \in H^\infty$ if and only if
$$\prod_{\nu \neq \mu} \left| \frac{a_\nu - a_\mu}{1 - a_\nu \bar{a}_\mu} \right| \geqslant \gamma > 0 \quad (\mu = 1, 2, \ldots). \tag{5}$$

The proofs of the above results are published in [2].

We shall now finally prove that if (5) holds, there exists an absolutely convergent interpolation series. The proof of this is due to the late Pehr Beurling.

Consider the variational problem
$$M = \sup_G \sum_1^n \lambda_\nu |G(a_\nu)| \quad (\|G\|_1 = 1).$$
Clearly we can assume $G \neq 0$ and put $G = e^{u+iv}$. A variation shows that the extremal function G has to satisfy the relation
$$e^{u(z)} = \frac{1}{M} \sum_{\nu=1}^n \lambda_\nu e^{u(a_\nu)} \frac{1 - |a_\nu|^2}{|e^{i\theta} - a_\nu|^2} \quad (z = e^{i\theta}).$$

Let us now apply this to $\lambda_\nu = |A'(a_\nu)|^{-1}$. In this case
$$M = \sup_{|w_\nu| \leqslant 1} \inf_f \|f\|_\infty$$

by (2). We now form the functions

$$f_\nu(z) = \frac{A(z)}{A'(a_\nu)(z-a_\nu)} \cdot \frac{e^{u(a_\nu)}}{G(z)} \cdot \frac{1-|a_\nu|^2}{1-z\bar{a}_\nu}.$$

Clearly $|f_\nu(a_\nu)| = 1$, $f_\nu(a_\mu) = 0$, $\mu \neq \nu$. Furthermore if $z = e^{i\theta}$,

$$\sum_1^n |f_\nu(z)| = \sum_{\nu=1}^n \frac{1}{|A'(a_\nu)|} e^{u(a_\nu)} \frac{1-|a_\nu|^2}{|e^{i\theta}-a_\nu|^2} \cdot e^{-u(z)} \equiv M.$$

Since the left-hand side is subharmonic

$$\sum_1^n |f_\nu(z)| \leqslant M \quad (|z| \leqslant 1),$$

and $\sum_{\nu=1}^n w_\nu f_\nu(z)$ solves the interpolation $f(a_\nu) = w_\nu$ and $\|f\| \leqslant M$.

It is easy to let $n \to \infty$ and we obtain an absolutely convergent interpolation series with the best possible bound M.

REFERENCES

[1]. CARLESON, L., An interpolation problem for bounded analytic functions. *Amer. J. Math.*, 80 (1958), 921–930.

[2]. —— Interpolations by bounded analytic functions and the Corona problem. *Ann. Math.*, 76 (1962), 547–559.

LES CÔNES CONVEXES FAIBLEMENT COMPLETS DANS L'ANALYSE

Par GUSTAVE CHOQUET

Les sous-ensembles convexes des espaces vectoriels topologiques localement convexes séparés (en abrégé e.l.c.) jouent un rôle de plus en plus important dans l'Analyse; je n'ai pas l'intention d'en étudier ici tous les aspects; je me bornerai à l'étude des cônes convexes localement compacts (ou, ce qui revient au même, des ensembles convexes compacts), et à l'étude des cônes convexes faiblement complets, qui constituent une généralisation fort utile des premiers.

Je m'attacherai plus particulièrement à l'étude des frontières de ces cônes et à la représentation de leurs points comme résultante de mesures portées par ces frontières. J'indiquerai également les relations entre cette théorie et celle des frontières associées à un espace vectoriel ou à une algèbre de fonctions continues sur un espace compact ou localement compact.

I. Ensembles convexes compacts

Désignons par X un sous-ensemble convexe compact d'un e.l.c. E, et par $\mathcal{E}$ l'ensemble des points extrémaux de X. Le théorème de Krein et Milman nous montre que tout point de X est limite de barycentres de mesures positives discrètes portées par $\mathcal{E}$. Lorsque X est métrisable, j'ai montré [9] que tout point de X est exactement le barycentre d'une mesure de Radon positive portée par $\mathcal{E}$, et j'ai caractérisé les cas où cette mesure est unique.

Ces résultats ont reçu ces dernières années des améliorations successives :

E. Bishop et K. de Leeuw [6] ont abordé le théorème d'existence dans le cas non-métrisable et montré l'intérêt d'introduire un ordre sur l'ensemble des mesures positives sur X.

H. Bauer [2 et 3] a étudié la frontière associée à une famille de fonctions et introduit l'idée de caractériser les cas d'unicité par des propriétés fonctionnelles.

M. Hervé [19], puis F. F. Bonsall [8] ont donné des démonstrations d'existence très simples dans le cas métrisable.

G. Mokobodzki [25] a caractérisé les mesures maximales dans le cas général.

P. A. Meyer [22] et G. Choquet [12] ont donné de nouvelles caractérisations du cas d'unicité, et un exposé d'ensemble de ces questions [16] d'où est tirée la partie I du présent travail.

On va voir que, grâce au théorème de Hahn-Banach et à la notion de mesure maximale, le théorème d'existence a maintenant une démonstration fort simple; on remarquera aussi que la notion de point extrémal fait son apparition très tard dans la théorie; cette remarque ne prendra d'ailleurs toute sa valeur que dans l'étude des cônes faiblement complets, lorsque nous aurons rencontré des exemples de tels cônes dépourvus de génératrices extrémales.

I.1. *Définitions et notations*

S désigne le sous-cône convexe de $\mathcal{C}(X)$ constitué par les fonctions convexes; il est total dans $\mathcal{C}(X)$.

$A = S \cap (-S)$ est l'ensemble des fonctions affines continues sur X.

$\mathcal{M}^+$ (resp. $\mathcal{M}^1$) est le cône convexe des mesures de Radon $\geqslant 0$ sur X (resp. $\geqslant 0$ et de norme 1), muni de la topologie faible.

Pour toute $\mu \in \mathcal{M}^+$ on désigne par $r(\mu)$ sa résultante $\int x d\mu(x)$.

Pour toute $f \in \mathcal{C}(X)$, on désigne par $\hat{f}$ la fonction concave s. c. s. sur X définie par.

$$\hat{f} = \inf.\ g \quad (\text{où } g \in -S,\ g \geqslant f).$$

L'application $f \to \hat{f}$ est évidemment croissante et sous-linéaire ($\widehat{f+g} \leqslant \hat{f} + \hat{g}$ et $\widehat{\lambda f} = \lambda \hat{f}$ si $\lambda \geqslant 0$); et pour toute $f \in -S$, on a $f = \hat{f}$. On montre que, pour tout $x \in X$, on a:

$$\hat{f}(x) = \sup.\ \mu(f) \quad (\text{où } \mu \in \mathcal{M}^1 \text{ et } r(\mu) = x).$$

On a la même relation pour les μ discrètes.

Pour toute $\mu \in \mathcal{M}^+$, on note $\hat{\mu}$ la fonction définie sur $\mathcal{C}(X)$ par $\hat{\mu}(f) = \mu(\hat{f})$; elle est croissante et sous-linéaire.

On appelle *ensemble bordant* X toute partie de X de la forme $\{x : f(x) = \hat{f}(x)\}$, où $f \in S$. Chacun de ces ensembles est un G_δ dont le complémentaire (dans X) est convexe. Ces ensembles vont jouer un rôle important. Leur intersection est l'ensemble $\mathcal{E}$ des points extrémaux de X.

I.2. *Mesures maximales et théorème d'existence*

Nous définissons dans $\mathcal{M}^+$ un ordre, noté $\prec$, en posant:

$$(\mu \prec \nu) \text{ si l'on a } \mu(f) \leqslant \nu(f) \text{ pour toute } f \in S.$$

Une mesure maximale pour cet ordre sera simplement dite *maximale*.

La relation $\mu \prec \nu$ entraîne $\mu(f) = \nu(f)$ pour toute $f \in A = S \cap (-S)$; donc μ et ν ont même résultante et même norme.

Par exemple, pour toute $\mu \in \mathcal{M}^1$, de résultante x, on a $\varepsilon_x \prec \mu$; cet exemple suffit à suggérer que $\mu \prec \nu$ signifie, en gros, que ν est plus dispersée, plus proche du bord de X, que μ; d'où l'intérêt des mesures maximales, qu'on espère portées par $\mathcal{E}$, ou du moins par chacun des ensembles bordant X.

THÉORÈME 1. *Toute $\mu \in \mathcal{M}^+$ est majorée, au sens de l'ordre $\prec$, par une mesure maximale.*

En vertu du théorème de Zorn, c'est une conséquence du fait, facile à établir, que $\mathcal{M}^+$ est inductif pour l'ordre $\prec$.

LEMME 2. *Soit $\mu \in \mathcal{M}^+$, et soit ν une forme linéaire sur $\mathcal{C}(x)$;*
$(\nu \leqslant \hat{\mu}$ *sur* $\mathcal{C}(X)) \Leftrightarrow (\nu \in \mathcal{M}^+$ *et* $\mu \prec \nu)$.

La vérification est immédiate.

THÉORÈME 3. *Soit $\mu \in \mathcal{M}^+$; les énoncés suivants sont équivalents.*

1. *μ est maximale*
2. *$\hat{\mu}$ est linéaire sur $\mathcal{C}(X)$*
3. *$\hat{\mu} = \mu$ sur $\mathcal{C}(X)$*
4. *$\hat{\mu} = \mu$ sur S*

Démonstration. On va montrer que $1 \Leftrightarrow 2$, et que $2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 2$.

La clef de la démonstration est la partie ci-dessous du théorème de Hahn-Banach :

Si p désigne une fonction numérique sous-linéaire sur un espace vectoriel V, dire qu'il existe une seule forme linéaire l sur V telle que $l \leqslant p$ équivaut à dire que p est linéaire (et alors évidemment $l = p$).

$1 \Leftrightarrow 2$. D'après le lemme 2, on a :

(μ maximale) $\Leftrightarrow$ (il existe une seule forme linéaire ν sur $\mathsf{C}(x)$ telle que $\nu \leqslant \hat{\mu}$).

D'après l'énoncé de Hahn-Banach, ceci équivaut aussi à dire que $\hat{\mu}$ est linéaire.

$2 \Rightarrow 3 \Rightarrow 4$. Si μ est linéaire, comme μ est aussi linéaire et que $\mu \leqslant \hat{\mu}$, on a $\mu = \hat{\mu}$.

$4 \Rightarrow 2$. On a toujours $\mu = \hat{\mu}$ sur $(-S)$; donc si $\mu = \hat{\mu}$ sur S, $\hat{\mu}$ est linéaire sur chaque droite de $S \cup (-S)$; comme de plus elle est sous-linéaire sur $(S - S)$, elle est linéaire sur $(S - S)$, donc aussi sur $\mathsf{C}(X)$.

COROLLAIRE 4. *Les énoncés suivants sont équivalents:*

(μ *maximale*) $\Leftrightarrow$ (μ *portée par chacun des ensembles bordant* X).

COROLLAIRE 5. *Tout $x \in X$ est résultante d'une $\mu \in \mathcal{M}^1$ portée par chacun des ensembles bordant X.*

COROLLAIRE 6. *L'ensemble M des mesures maximales est un sous-cône convexe de $\mathcal{M}^+$, stable par intégration, et héréditaire à gauche. Ce cône convexe M est réticulé pour son ordre propre.*

Ce sont des conséquences immédiates du théorème 3.

I.3. *Mesures maximales et points extrémaux*

Les démonstrations et résultats qui précèdent ne font pas intervenir l'ensemble $\mathcal{E}$ des points extrémaux. Nous nous proposons maintenant d'étudier jusqu'à quel point on peut dire que les mesures maximales sont portées par $\mathcal{E}$.

La situation n'est tout à fait simple que lorsque X est métrisable :

LEMME 7. *Lorsque X est métrisable, $\mathcal{E}$ est le plus petit ensemble bordant X.*

Cela résulte de ce qu'alors il existe une $f \in S$ qui est strictement convexe, ce qui entraîne que $\mathcal{E}$ = l'ensemble bordant $\{x : f(x) = \hat{f}(x)\}$.

PROPOSITION 8. *Lorsque X est métrisable, $\mathcal{E}$ est un G_δ; les mesures maximales sont celles portées par $\mathcal{E}$, et tout $x \in X$ est résultante d'une $\mu \in \mathcal{M}^1$ portée par $\mathcal{E}$.*

Les résultats du paragraphe précédent permettent aussi d'énoncer immédiatement la proposition suivante :

PROPOSITION 9. *Supposons à nouveau X quelconque, et soit $\mu \in \mathcal{M}^+$; alors,* (μ *portée par tout ouvert de X contenant $\mathcal{E}$*) $\Rightarrow$ (μ *maximale*).

La réciproque de cette proposition est en général inexacte. On a toutefois l'énoncé suivant :

PROPOSITION 10. (μ *maximale*) $\Rightarrow$ (μ *portée par tout ensemble K-souslinien contenant* $\mathcal{E}$).

Dans cet énoncé on appelle K-souslinienne toute partie de X obtenue à partir des fermés de X par l'opération-A (par exemple les ensembles K-boréliens, tels que les K_σ, $K_{\sigma\delta}$, $K_{\sigma\delta\sigma}$, etc.)

Notons ici qu'en particulier toute mesure maximale est portée par tout ouvert de Baire contenant $\mathcal{E}$; mais la réciproque est fausse : Plus précisément, il peut arriver qu'une $\mu \in \mathcal{M}^+$ soit portée par tout ensemble K-souslinien contenant $\mathcal{E}$, sans pour cela être maximale. D'où la valeur du critère de maximalité fourni par le corollaire 4.

I.4. *Théorème d'unicité*

Il est commode, dans ce qui suit, de considérer X comme la base d'un cône convexe C, c'est-à-dire que X est l'intersection de C avec un hyperplan affine fermé ne contenant pas O et rencontrant toutes les génératrices de C; ceci est réalisable quel que soit X : Il suffit d'identifier E à l'hyperplan affine $E \times \{1\}$ de $E \times R$ et de prendre pour C le cône engendré par X.

Il est immédiat que tous ces cônes C sont linéairement isomorphes; en particulièr si l'un d'eux est réticulé pour son ordre propre, tous le sont; on dit dans ce dernier cas que X est un *simplexe* (parce que lorsque X est de dimension finie, cette condition entraîne l'isomorphie de X avec les simplexes de la géométrie élémentaire : triangles, tétraèdres, etc.).

THÉORÈME 11 (*d'unicité*). *Pour tout convexe compact X, les propriétés suivantes sont équivalentes :*

1. *X est un simplexe*
2. *Pour toute $f \in S$, l'application $x \to \hat{f}(x)$ est affine dans X.*
3. *Pour toute μ maximale $\in \mathcal{M}^1$ et pour toute $f \in S$, on a $\hat{f}(x) = \mu(f)$, où $x = r(\mu)$.*
4. *L'application $f \to \hat{f}$ est linéaire sur S.*
5. *Tout $x \in X$ est résultante d'une mesure maximale unique de $\mathcal{M}^1$, notée μ_x.*

La démonstration se fait dans l'ordre naturel : $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 1$; on y utilise le fait que, pour toute fonction f affine s. c. s. sur X et toute $\mu \in \mathcal{M}^1$, on a $\mu(f) = f(x)$, où x est la résultante de μ.

Notons que l'application $x \to \mu_x$ de X dans $\mathcal{M}^1$ est de 1ère classe, donc universellement mesurable.

C'est évidemment l'équivalence des énoncés 1 et 5 qui est la plus utilisée dans les applications; plus précisément on utilise souvent l'énoncé suivant :

COROLLAIRE 12. *Si X est base d'un cône convexe réticulé C, et si $\mathcal{E}$ est un ensemble K-souslinien (par exemple si X est métrisable), tout $x \in C$ est résultante d'une μ unique de $\mathcal{M}^+$ portée par $\mathcal{E}$.*

Il est parfois très facile de montrer qu'un cône C est réticulé; voici quelques exemples de tels cas :

Le cône des fonctions harmoniques $\geqslant 0$ sur un espace de Green.

Le cône des mesures positives sur un espace compact, qui sont invariantes par une famille de diffusions continues et conservatives.

Le cône des mesures $\mu \geqslant 0$ sur un groupe abélien localement compact, satisfaisant à une relation $\alpha * \mu \leqslant \mu$, ou α est une mesure $\geqslant 0$ donnée.

Le fait qu'un cône C est réticulé est parfois plus caché; on peut alors tenter d'utiliser les critères 2, 3, 4, ou les critères 6, 7, 8 qui suivent :

6. *L'ensemble des dilatés positifs* $(k\,X+a)$ *de* X *(où* $k\in R^+$ *et* $a\in E$*) contenant* O *est stable par intersection finie.*

L'équivalence de 1 et 6 est de nature purement algébrique.

7. *Le cône convexe des fonctions numériques affines s. c. s. sur* X, *ordonné par* $(f_1\leqslant f_2$ *si* $f_1(x)\leqslant f_2(x)$ *pour tout* $x\in X)$ *est semi-réticulé supérieurement.*

Le critère suivant est directement équivalent au critère 2; on y désigne par $L(\mathcal{E})$ l'ensemble de traces sur $\mathcal{E}$ des fonctions affines continues sur E.

8. *Pour toute famille finie* (f_i) *d'éléments de* $L(\mathcal{E})$, *l'ensemble des* $f\in L(\mathcal{E})$ *qui sont* $\geqslant\sup\,(f_i)$ *est filtrante décroissante et a pour limite inférieure* $\sup\,(f_i)$.

On utilise souvent ce critère sous la forme affaiblie suivante :

Soit H un ensemble quelconque, K un espace topologique compact, et soit f une application de $H\times K$ dans R, telle que pour tout $x\in H$ l'application $f_x: y\to f(x,y)$ soit continue dans K, et telle que l'ensemble des f_x sépare K.

Soit X l'ensemble, muni de la topologie de la convergence simple, des fonctions numériques g sur H de la forme : $g(x)=\int f(x,y)\,d\mu(y)$, où $\mu\in\mathcal{M}^1(K)$.

Alors X est convexe compact; désignons par K' le sous-ensemble de X homéomorphe à K, constitué par les fonctions g de la forme $g(x)=f(x,a)$. On a alors l'équivalence très commode :

(X *est un simplexe et* $K'=\mathcal{E}(X)$) $\Leftrightarrow$ (*l'ensemble des* f_x *est total dans* $\mathcal{C}(K)$)

II. Les cônes convexes faiblement complets

II. 1. Malgré leur grand intérêt, les cônes convexes saillants localement compacts (dont l'étude se ramène à celle de leurs bases affines, qui sont des convexes compacts) ne constituent qu'une petite partie d'une classe plus vaste, celle des cônes convexes faiblement complets; et les exemples qui suivent montrent que la plupart des cônes faiblement complets que l'on rencontre en analyse ne sont pas localement compacts.

Exemples. 1. Le cône des fonctions convexes positives sur un ouvert-convexe de R^n.

2. Le cône des fonctions $\geqslant 0$ ainsi que toutes leurs dérivées sur un intervalle ouvert de R.

3. Le cône des mesures de Radon $\geqslant 0$ sur un espace localement compact, et ses sous-cônes convexes fermés : Par exemple le cône des mesures $\mu\geqslant 0$ satisfaisant à la relation $\alpha*\mu\leqslant\mu$ sur un groupe localement compact (α est une mesure $\geqslant 0$ donnée).

4. Le cône des capacités alternées d'ordre infini sur l'ensemble $\mathcal{K}(X)$ des compacts d'un espace localement compact X.

5. Le cône des fonctions à valeurs complexes, de type positif sur le carré E^2 d'un ensemble quelconque E.

6. Plus généralement, soit A une partie d'un espace vectoriel E, qui engendre E. Le cône des formes linéaires sur V qui sont positives sur A est saillant et faiblement complet dans le dual de V.

Nous n'avons pas précisé dans ces exemples la topologie à mettre sur les cônes étudiés; elle est en général évidente.

L'interêt de la classe des cônes convexes faiblement complets tient à

plusieurs facteurs : Sa grande généralité, sa stabilité par limite projective (autrement dit par produit et par hérédité), la commodité de la topologie faible, et le grand nombre de propriétés utiles qu'elle possède malgré sa généralité (voir [11]).

II.2. *Structure des convexes faiblement complets*

On dira que la topologie d'un espace vectoriel topologique E (sur R) est *faible* si elle est séparée et identique à la topologie affaiblie de E. Cette propriété se transmet aux produits, sous-espaces, quotients par un sous-espace fermé, complétés. On sait que tout E faible complet est isomorphe à un produit R^I et que dans un tel espace, tout sous-espace fermé a un supplémentaire topologique.

On dira qu'une partie X d'un e.v.t. E est *faiblement complète* si la topologie affaiblie de E est séparée, et si X est complet pour la structure uniforme associée (autrement dit si X est fermé dans le complété faible de E).

THÉORÈME 13. *Pour tout convexe complet X d'un e. v. t. faible E, il existe une isomorphie φ de E dans un produit $R^I \times R^J$ telle qu'un translaté convenable de $\varphi(X)$ s'écrive $R^I \times Y$ où Y est un convexe fermé de R^J_+.*

Cet énoncé est fondamental et tous les corollaires qui suivent en découlent simplement; il ramène l'étude des convexes faiblement complets à l'étude de ceux qui sont *saillants*, c'est-à-dire ne contiennent aucune droite; on désignera par $\mathcal{S}$ la classe de ces derniers.

Nous venons de voir que tout $Y \in \mathcal{S}$ est plongeable dans un R^J_+. La classe $\mathcal{S}$ est la plus petite classe, stable par limite projective, et contenant les convexes fermés saillants des espaces R^n de dimension finie.

CAS PARTICULIER. *Tout $X \in \mathcal{S}$ dont la structure uniforme est métrisable est isomorphe à un convexe fermé de R^N_+ (où N est l'ensemble des entiers naturels).*

Voici maintenant quelques conséquences du théorème 13.

COROLLAIRE 14. *Soit X un convexe faiblement complet d'un e. v. t. E.*
(X est saillant)$\Leftrightarrow$(Toute forme affine continue sur E est différence de deux formes affines continues qui sont $\geqslant 0$ sur X.

COROLLAIRE 15. *Pour tout $X \in \mathcal{S}$ (où $X \subset E$) et tout $a \in E$, $X \cap (a - X)$ est compact.*

Autrement dit, tout X symétrique de $\mathcal{S}$ est compact. Plus généralement,

COROLLAIRE 16. *Soit $X \in \mathcal{S}$, et soient (x_i) et (y_i) $(i \in I)$ deux familles de points de X. Si $(x_i + y_i)$ converge suivant un ultrafiltre sur I, il en est de même de (x_i) et (y_i).*

COROLLAIRE 17. *Pour tous fermés A, B d'un $X \in \mathcal{S}$, $A + B$ est fermé.*

Cône asymptote d'un convexe fermé. Soit X un convexe fermé d'un e. v. t. E; pour tout $a \in X$, $\bigcap_{\lambda > 0} \lambda(X - a)$ est un cône convexe fermé qui ne dépend pas de a; c'est le *cône asymptote* de X, qu'on notera $\alpha(X)$. Il est identique à l'ensemble des valeurs d'adhérence de λx $(\lambda > 0,\ x \in X)$ lorsque $\lambda \to 0$.

Dire que $\alpha(X) = \{0\}$ équivaut à dire que X ne contient aucune demi-droite; si de plus X est alors localement compact, X est compact. Mais il existe des convexes métrisables $X \in \mathcal{S}$, non compacts, et tels que $\alpha(X) = \{0\}$.

COROLLAIRE 18. *Si* $X, Y \in \mathcal{S}$ *et* $(X+Y) \in \mathcal{S}$, *on a* $\alpha(X+Y) = \alpha(X) + \alpha(Y)$.

COROLLAIRE 19. *Soient* $X_i \subset Y$ $(i=1, 2, \ldots, n)$ *avec* X_i *et* $Y \in \mathcal{S}$. *Soit* A *l'enveloppe convexe des* X_i; *alors* :

$$\bar{A} = A + \sum \alpha(X_i) = \textit{enveloppe convexe des } (X_i + \alpha(X_j)); \ \alpha(\bar{A}) = \sum \alpha(X_i)$$

Voici une conséquence facile, mais importante, du corollaire 19:

COROLLAIRE 20. *Tout point extrémal d'un* $X \in \mathcal{S}$ *est extrémal fort (c'est-à-dire a dans* X *une base de voisinages de la forme* $f < 0$, *où* f *est une forme affine continue sur* E.

On a un énoncé analogue pour les génératrices extrémales des cônes $X \in \mathcal{S}$.

II.3. *Eléments extrémaux des cônes convexes saillants faiblement complets*

Il serait extrémement désirable de disposer, pour les cônes $X \in \mathcal{S}$, d'un théorème analogue à celui de Krein et Milman et, éventuellement, d'un théorème de représentation intégrale analogue à celui établi pour les cônes localement compacts.

Dès qu'on aborde ce problème, se présente une première difficulté : C'est qu'un cône $X \in \mathcal{S}$ n'a pas nécessairement de base affine; c'est le cas par exemple pour le cône $\mathcal{M}^+(R)$ des mesures de Radon positives sur la droite réelle. Nous sortirons de cette difficulté grâce à la notion de *chapeau* qui va se substituer dans certains cas à celle de base affine :

Une difficulté plus sérieuse est qu'il existe effectivement des cônes $X \in \mathcal{S}$ ne possédant aucune génératrice extrémale [14]:

Exemple de cône $X \in \mathcal{S}$ *sans génératrices extrémales.* Désignons par $f_{a,\alpha}$ l'application de [0, 1] dans [0, ∞] definie par

$$x \to 1/|x-a|^{\alpha}, \text{ où } a \in [0, 1] \text{ et } 0 \leqslant \alpha < 1.$$

et désignons par X le cône convexe des mesures de Radon positives μ sur [0, 1] telles que toute $f_{a,\alpha}$ soit intégrable.

Ce cône n'a aucune génératrice extrémale; et il est faiblement complet pour la topologie faible associée aux formes linéaires sur $(X-X)$ de la forme $\mu \to \mu(f_{a,\alpha})$; il est d'ailleurs aussi réticulé.

L'existence d'un tel exemple nous conduit, d'une part à chercher des classes aussi vastes que possible de cônes de $\mathcal{S}$ qui soient l'enveloppe convexe fermée de l'ensemble de leurs génératrices extrémales, et d'autre part à chercher une forme des théorèmes de représentation intégrale valable pour tous les cônes de $\mathcal{S}$.

Les chapeaux des cônes convexes.

Soit C un cône convexe d'un e. v.t. E. On appelle *chapeau* de C toute partie convexe compacte A de C dont le complémentaire dans C soit aussi convexe.

Il est immédiat que A contient alors 0, et que tout point extrémal de A distinct de 0 appartient à une génératrice extrémale de C. On en déduit aussitôt l'énoncé suivant :

PROPOSITION 21. *Dans un e. l. c.* E, *tout point* a *d'un cône convexe* C *qui appartient à un chapeau* A *de* C *appartient à l'enveloppe convexe fermée des génératrices extrémales de* C.

Si ce chapeau A est métrisable, a est résultante d'une mesure de Radon positive portée par les génératrices extrémales de C, et de support compact.

Si C est réticulé, tout chapeau de C est un simplexe.

Nous allons maintenant déterminer des classes étendues de cônes de $\mathcal{S}$ suffisamment riches en chapeaux.

LEMME 22. 1) *Si A est un chapeau d'un cône convexe C, et si C' est un sous-cône convexe fermé de C, $A \cap C'$ est un chapeau de C'.*

2) *Si A_n est un chapeau de C_n ($n \in N$), il existe un chapeau A de ΠC_n qui contient ΠA_n.*

En voici une conséquence immédiate:

THÉORÈME 23. *Si C_n est réunion de ses chapeaux, et si $C = \varprojlim C_n$ (limite projective), C est aussi réunion de ses chapeaux.*

C'est en particulier le cas des cônes de $\mathcal{S}$ qui sont limite projective d'une suite de cônes localement compacts, par exemple tout sous-cône fermé du cône des mesures de Radon $\geqslant 0$ sur un espace localement compact dénombrable à l'infini, ou tout cône métrisable $\in \mathcal{S}$.

COROLLAIRE 24. *Pour tout convexe $X \in \mathcal{S}$ qui est localement compact ou limite projective d'une suite d'éléments de $\mathcal{S}$ localement compacts, on a:*

$X = X' + \alpha(X)$ (où X' désigne l'enveloppe convexe fermée de l'ensemble des points extrémaux de X).

En particulier cette relation entraîne qu'un tel X non-vide a au moins un point extrémal et que, s'il n'en contient qu'un, c'est un cône ayant ce point pour sommet. Si $\alpha(X) = \{0\}$, on a $X = X'$, c'est-à-dire que X est l'enveloppe convexe fermée de ses points extrémaux.

Remarque. Si un cône $X \in \mathcal{S}$ est limite projective d'une suite de cônes localement compacts, ce qui précède montre que tout sous-cône fermé de X est réunion de ses chapeaux. Mais il existe d'autres cônes de $\mathcal{S}$ qui possèdent cette propriété d'hérédité; en voici un exemple intéressant : Soit E un espace vectoriel de base algébrique infinie dénombrable B; on le munit de la plus fine de toutes les topologies faibles; le cône X engendré par B est le cône cherché; il a en outre la propriété curieuse de ne pas être un espace de Baire bien qu'étant un convexe fermé d'un espace R^I.

III. Mesures coniques et cônes faiblement complets

Dans la premiére partie de cet exposé, nous avons ramené l'étude des représentations intégrales dans un cône localement compact de $\mathcal{S}$ à la même étude sur une de ses bases affines; on peut objecter à ce procédé qu'il n'est pas intrinsèque puisqu'il fait jouer un rôle privilégié à une base affine particulière.

D'autre part ce procédé ne peut pas s'étendre au cas des cônes faiblement complets quelconques puisque en général ceux-ci n'ont pas de base affine.

Ces deux faits rendent souhaitable la création d'êtres mathématiques qui remplacent les mesures portées par une base d'un cône, et qui aient un caractère intrinsèque. Nous allons montrer que les mesures coniques

répondent parfaitement à la question. Elles nous permettront de construire de façon simple une théorie analogue à celle des mesures maximales sur les convexes compacts, sans utiliser à aucun instant la notion de génératrice extrémale [13 et 14].

III.1. *Définition et étude des mesures coniques*

Soient E, F deux e. v. t. sur R. Une application continue f de E dans F est dite *linéaire par morceaux* s'il existe un recouvrement fini de E par des *cônes polyédraux* E_i (intersection finie de demi-espaces fermés) sur chacun desquels f soit linéaire; on désigne par $h(E,F)$ l'ensemble de ces f; c'est un espace vectoriel qui contient $L(E,F)$ et l'on a:

$$h(F,G)\circ h(E,F)\subset h(E,G).$$

En particulier $h(E,R)$ est identique à l'ensemble des différences $f=f_1-f_2$, où f_1 et f_2 sont de la forme sup $(l_1,l_2,\ldots,l_n)$, où $l_i\in L(E,R)$; l'espace $h(E,R)$ est stable par l'opération $f\to|f|$.

Nous supposerons désormais, pour simplifier, que les e. v. t. utilisés, E, F, etc. sont faibles; on notera E' le dual topologique de E.

DÉFINITION. *On appelle mesure conique sur E toute forme linéaire μ sur $h(E,R)$ telle que $\mu(f)\geqslant 0$ pour toute $f\geqslant 0$.*

Le cône convexe $\mathcal{M}_c^+(E)$ des mesures coniques sur E est complètement réticulé et faiblement complet pour la topologie faible sur le dual algébrique de $h(E,R)$ associée à la forme bilinéaire $(\mu,f)\to\mu(f)$.

Exemples. 1) Pour tout $x\in E$, la forme linéaire $f\to f(x)$ est une mesure conique que l'on notera ε_x. On appellera *mesure discrète* toute combinaison liniéaire finie à coefficients positifs de telles mesures ε_x.

2) Si K est un compact de E, et ν une mesure de Radon $\geqslant 0$ sur K, on définit une mesure conique μ en posant: $\mu(f)=\nu(f_k)$ (où f_k désigne la restriction à K d'une $f\in h(E,R)$.

On dit d'une telle μ qu'elle est *localisable* sur K. Il est inexact que toute mesure conique soit localisable sur un compact.

On dit qu'une mesure conique μ est *portée* par un cône fermé X de E si $\mu(f)\geqslant 0$ pour toute f qui est $\geqslant 0$ sur X; le cône convexe $\mathcal{M}_c^+(X)$ des mesures coniques portées par X est fermé dans $\mathcal{M}_c^+(E)$ et complètement réticulé.

On appelle *support* de μ la réunion $S\mu$ des demi-droites δ de E telles que la condition ($f\geqslant 0$ partout, et $f>0$ sur δ ouverte) entraîne $\mu(f)>0$; l'ensemble $S\mu$ est l'intersection des cônes fermés qui portent μ, il ne porte pas nécessairement μ et peut même être vide sans que $\mu=0$.

Toute mesure conique sur E se prolonge de façon évidente en une mesure conique sur le complété faible $\hat{E}$ de E. Pour toute mesure conique μ, il existe un point unique x de E tel que $\mu(f)=f(x)$ pour toute $f\in E'$; on appelle ce point la *résultante* de μ et on le note $r(\mu)$; lorsque μ est portée par un cône convexe faiblement complet A de E, $r(\mu)$ appartient à A.

Pour toute $\varphi\in h(E,F)$, la relation $\varphi(\mu)\,(f)=\mu(f\circ\varphi)$ définit l'image $\varphi(\mu)$ de la mesure conique μ par φ; si μ est portée par A, $\varphi(\mu)$ est portée par $\overline{\varphi(A)}$.

Le lemme suivant permet souvent de démontrer des propriétés des mesures coniques à partir de propriétés évidentes des mesures discrètes :

LEMME 25. *Soit μ une mesure conique sur E, portée par un cône fermé A.*
1. *μ est limite de mesures discrètes portées par A.*
2. *Lorsque A est convexe et faiblement complet, μ est limite de mesures discrètes portées par A et de même résultante que μ.*

Voici maintenant quelques résultats concernant la localisation des mesures coniques.

PROPOSITION 26. *Soit X un cône convexe faiblement complet de E, métrisable ou plus généralement limite projective d'une suite de cônes localement compacts.*
Alors toute mesure conique μ portée par X est localisable sur un compact de X (ce qui entraîne que μ soit portée par son support $S\mu$).
Si de plus X est saillant, μ est localisable sur tout chapeau de X contenant la résultante de μ.

On peut déduire de là que toute mesure conique μ portée par un cône convexe faiblement complet X est une intégrale de Daniell, en ce sens que que pour toute suite (f_n) d'éléments de $h(E,R)$ qui sur X tend vers O en décroissant, on a : $\lim \mu(f_n)=0$
Pour terminer cette esquisse, soulignons l'analogie entre les mesures coniques et les mesures étudiées par Gel'fand-Vilenkin [17] sous le nom de mesures cylindriques :
Si dans notre définition des mesures coniques on remplaçait l'espace $h(E,R)$ par l'espace des fonctions continues *affines* par morceaux et bornées dans E, on obtiendrait une notion équivalente à celle de mesure cylindrique.

III.2. *Mesures coniques maximales*

Soit X un cône de $\mathcal{S}$; désignons par S le sous-cône convexe de $h(E,R)$ constitué par les fonctions dont la restriction à X est convexe.
Nous définissons sur $\mathcal{M}_c^+(X)$ une relation d'ordre, notée $\prec$, en posant $(\mu\prec\nu)$ si $\mu(f)\leqslant\nu(f)$ pour toute $f\in S$.
La relation $\mu\prec\nu$ entraîne évidemment $\mu(f)=\nu(f)$ pour toute $f\in E'$, donc μ et ν ont même résultante.
Par exemple, pour toute μ de résultante x, on a $\varepsilon_x\prec\mu$.
Si une $\mu\in\mathcal{M}_c^+(X)$ est maximale pour l'ordre $\prec$, on dira simplement qu'elle est maximale.
Dès lors la théorie se déroule exactement comme celle exposée dans la 1ière partie pour les convexes compacts, avec des démonstrations au moins aussi simples. Introduisons d'abord la terminologie nécessaire :
Pour toute $\mu\in\mathcal{M}_c^+(X)$, $\hat{\mu}$ est l'application de $h(E,R)$ dans R définie par :

$$\hat{\mu}(f)=\inf\mu(g) \quad \text{pour les} \quad g\in -S \text{ qui sont } \geqslant f \text{ sur } X.$$

THÉORÈME 27. *Toute $\mu\in\mathcal{M}_c^+(X)$ est majorée, au sens de l'ordre $\prec$, par une mesure maximale.*

THÉORÈME 28. *Soit $\mu\in\mathcal{M}_c^+(X)$; les énoncés suivants sont équivalents :*

1. *μ est maximale*
2. *$\hat{\mu}$ est linéaire sur $h(E,R)$*
3. *$\mu=\hat{\mu}$ sur $h(E,R)$*
4. *$\mu=\hat{\mu}$ sur S*

COROLLAIRE 29. *Tout $x \in X$ est résultante d'une $\mu \in \mathcal{M}_c^+(X)$ « portée par le bord de X » en ce sens que $\hat{\mu}(f) = \mu(f)$ pour toute $f \in S$.*

COROLLAIRE 30. *L'ensemble des mesures maximales est un sous-cône convexe de $\mathcal{M}_c^+(X)$, stable par intégration et héréditaire à gauche. Ce cône est réticulé pour son ordre propre.*

Lors de l'étude des cônes faiblement complets nous avons construit un exemple de cône $X \in \mathcal{S}$ n'ayant aucune génératrice extrémale; on peut montrer que toute mesure maximale sur ce cône a un support vide.

Nous ne pouvons donc pas espérer, dans le cas général, caractériser les mesures maximales en termes de support et d'éléments extrémaux; en somme nous retrouvons, mais poussée à l'extrême, une difficulté qui se manifestait déjà partiellement dans les convexes compacts non métrisables :

THÉORÈME 31. *Soit $\mu \in \mathcal{M}_c^+(X)$ et soit K un chapeau de X contenant la résultante de μ.*
(μ est maximale) $\Leftrightarrow$ (Il existe une localisation de μ sur K qui est maximale sur K)

Ce théorème est particulièrement utile, en vertu du théorème 23 et de la proposition 26, lorsque X est limite projective d'une suite de cônes localement compacts.

En particulier, si X est métrisable, dire que μ est maximale equivaut à dire que μ est portée, en un sens facile à préciser, par l'ensemble des génératrices extrémales de X.

III.3. *Critères d'unicité*

Le théorème d'unicité s'énonce et se démontre, lui aussi de façon presque identique au théorème d'unicité relatif aux convexes compacts. Donnons d'abord une definition :

Pour toute $f \in h(E, R)$ on définit ainsi l'application $\hat{f}$ de X dans R :

$$\hat{f}(x) = \inf l(x) \quad (\text{où } l \in E' \text{ et } l \geqslant f \text{ sur } X)$$

THÉORÈME 32 (*d'unicité*). *Pour tout cône $X \in \mathcal{S}$, les propriétés suivantes sont équivalentes :*

1. *X est réticulé pour son ordre propre.*
2. *Pour toute $f \in S$, l'application $x \to \hat{f}(x)$ est linéaire sur X.*
3. *Pour toute μ maximale $\in \mathcal{M}_c^+(X)$ et pour toute $f \in S$, on a : $\hat{f}(x) = \mu(f)$, où $x = r(\mu)$*
4. *L'application $f \to \hat{f}$ est linéaire sur S.*
5. *Tout $x \in X$ est résultante d'une μ maximale unique $\in \mathcal{M}_c^+(X)$.*

Le fait que, lorsque X est réticulé, tout $x \in X$ soit résultante d'une mesure conique maximale unique montre bien que les mesures coniques constituent le bon outil pour la représentation intégrale des points de X, malgré la possibilité que toutes ces mesures maximales aient un support vide.

Il n'est pas exclu cependant que l'on puisse encore améliorer cette étude en introduisant, à côté des éléments extrémaux classiques, des éléments-frontière idéaux sur lesquels on pourrait localiser les mesures maximales. C'est du moins ce qui ressort de l'étude de l'exemple du paragraphe II 3.

IV. Frontières associées à un espace vectoriel de fonctions continues

De nombreux résultats concernant les frontières associées à un espace vectoriel ou à une algèbre de fonctions continues à valeurs complexes sur un espace compact X ont été obtenus ces dernières années par R. Arens [1], H. Bauer [2 et 3], E. Bishop [4, 5, 6], K. de Leeuw [6], I. M. Singer [1].

On obtient un exposé plus simple et plus général, d'une part en utilisant des fonctions à valeurs dans un e.l.c. sur R, éventuellement normé; d'autre part en utilisant le plus possible les propriétés des convexes compacts. En outre, cette méthode permet d'étendre certains des résultats au cas où X est localement compact [15].

IV.1. *Cas de X compact*

On se propose l'étude des structures (X, E, B), où X est un espace topologique compact, E un e. l. c., et B un sous-espace vectoriel de $C(X, E)$ qui sépare X.

Si $\mu, \nu \in \mathfrak{M}^1(X)$, on dit que $\mu \sim \nu$ si pour toute $f \in B$ on a $\mu(f) = \nu(f)$, autrement dit si pour toute $f \in B$ et toute $l \in E'$, on a $\mu(l \circ f) = \nu(l \circ f)$.

Définition. *On appelle frontière fine de la structure (X, E, B) l'ensemble $\mathcal{E}(B)$ des $x \in X$ tels que ε_x soit la seule $\mu \in \mathfrak{M}^1(X)$ verifiant $\mu \sim \varepsilon_x$.*

Si nous notons $\tilde{B}$ le sous-espace $E' \circ B$ de $C(X, R)$ constitué par les fonctions de la forme $l \circ f$, où $l \in E'$ et $f \in B$, il est immédiat que $\mathcal{E}(B) = \mathcal{E}(\tilde{B})$.

Désignons par φ la bijection canonique de X dans $R^{\tilde{B}}$ définie par :

$$x \to (\text{le point de coordonnées } g(x)), \quad \text{où } g \in \tilde{B}.$$

Ce plongement φ de X identifie B à l'ensemble des traces sur $\varphi(X)$ des formes linéaires continues sur $R^{\tilde{B}}$ et $\varphi(\mathcal{E}(B))$ n'est autre que l'ensemble des points extrémaux de l'enveloppe convexe fermée (elle est compacte) de $\varphi(X)$.

Il devient alors évident que toute $g \in \tilde{B}$ atteint sa borne supérieure sur $\mathcal{E}(B)$, que $\overline{\mathcal{E}(B)}$ est la frontière de Šilov associée à $\tilde{B}$, et que $\mathcal{E}(B)$ est un G_δ lorsque X est métrisable. De même tous les théorèmes concernant les mesures maximales obtenus dans la partie I de cet exposé se transposent aussitôt aux structures (X, E, B).

Lemme 33. *Pour toute fonction numérique convexe continue p sur E et pour toute $f \in B$, $p(f)$ atteint sa borne supérieure sur $\mathcal{E}(B)$.*

Théorème 34. *Supposons E normable et muni d'une norme p differentiable (hors de 0); et supposons que B contienne les constantes (éléments de E).*

1. *Alors tout point $x \in \mathcal{E}(B)$ a dans X une base de voisinages de la forme $\{y : p(f(y)) > k\}$, où $f \in B$ et $k \in R$.*
2. *$\overline{\mathcal{E}(B)}$ est la frontière de Šilov pour la famille $\{p(f)\}_{f \in B}$.*

Ce théorème s'applique en particulier au cas où E est le corps C des complexes, et p la norme classique $z \to |z|$.

PROPOSITION 35. *Faisons les mêmes hypothèses sur E, B et p, et soit $x \in X$. Alors les énoncés suivants sont équivalents :*

1. $x \in \mathcal{E}(B)$
2. *Pour tout $\varepsilon > 0$ et pour tout ouvert ω de X contenant x, il existe $f \in B$ telle que :*

$$M - p(f(x)) < \varepsilon(p(f(x) - N),$$

où M (resp. N) $= \sup p(f)$ *sur X (resp. sur $(X \dot{-} \omega)$)*

3. *Même énoncé que* (2) *mais en remplaçant « pour tout ε » par « il existe ε ».*

Cette proposition permet aisément de montrer que si E est une algèbre de Banach sur R, de norme différentiable et telle que pour tout $a \in E$ et tout entier n, $\|a^n\| = \|a\|^n$, et si B est une algèbre de Banach contenant les constantes, alors pour tout $a \in \mathcal{E}(B)$ ayant dans X une base dénombrable de voisinages, il existe $f \in B$ telle que $\|f(x)\| < \|f(a)\|$ pour tout $x \neq a$.

IV.2. *Cas de X localement compact*

Lorsqu'on veut étendre les résultats qui précèdent au cas ou X est un espace topologique localement compact, on se heurte à une difficulté concrétisée par l'énoncé suivant :

PROPOSITION 36. *Il existe dans R^I_+ (où I a la puissance du continu) un sous-ensemble fermé localement compact X à base dénombrable, dont l'enveloppe convexe fermée n'a aucun point extrémal.*

Si l'on prend pour B l'espace vectoriel des traces, sur cet ensemble X, des formes linéaires continues sur R^I, toute frontière raisonnable que l'on pourrait associer à B se trouve vide, bien que $B = B_+ - B_+$.

On peut néanmoins retrouver certains énoncés antérieurs, à condition de les reformuler convenablement :

Soit X un espace topologique localement compact, et soit B un sous-espace vectoriel de $\mathcal{C}(X, R)$ qui sépare X, et tel que $B = \underline{B} - \underline{B}$, ou $\underline{B}$ désigne l'ensemble des $f \in B$ dont la borne inférieure sur X est finie.

On appelle *frontière fine* associée à B l'ensemble $\mathcal{E}(B)$ des points $x \in X$ ayant la propriété suivante :

Pour tout filtre $\mathcal{F}$ sur $\mathcal{M}^1(X)$ tel que, pour tout $f \in \underline{B}$ on ait :

$$f(x) = \lim_{\mathcal{F}} \mu(f), \ \mathcal{F} \text{ converge vaguement vers } \varepsilon_x.$$

PROPOSITION 37. *Si φ désigne l'injection canonique de X dans $R^{\underline{B}}$, $\varphi(\mathcal{E}(B))$ est identique à l'ensemble des points de $\varphi(X)$ qui sont extrémaux sur l'enveloppe convexe fermée de $\varphi(X)$.*

PROPOSITION 38. *Désignons par D l'ensemble des $f \in B$ pour lesquelles il existe $k \in R$ tel que l'ensemble $\{x : f(x) > k\}$ soit non-vide et relativement compact dans X.*

Alors toute $f \in D$ atteint sa borne supérieure sur $\mathcal{E}(B)$, et $\overline{\mathcal{E}(B)}$ est frontière de Šilov pour D.

BIBLIOGRAPHIE

[1]. ARENS, R. & SINGER, I. M., Function values as boundary integrals. *Proc. Amer. Math. Soc.*, 5 (1954), 735–745.

[2]. BAUER, H., Minimalstellen von Funktionen und Extremal punkte. I et II. *Arch. Math.*, 9 (1958), 389–393 et 11 (1960), 200–205.

[3]. BAUER, H., Šilovscher Rand und Dirichletsches Problem. *Ann. Inst. Fourier, Grenoble*, 11 (1961), 89–136.

[4]. BISHOP, E., A minimal boundary for function algebras. *Pacific J. Math.*, 9 (1959), 629–642.

[5]. —— Some theorems concerning function algebras *Bull. Amer. Math.-Soc.*, 85 (1959), 77–78.

[6]. BISHOP, E. & DE LEEUW, K., The representation of linear functionals by measures on sets of extreme points. *Ann. Inst. Fourier, Grenoble*, 9 (1959), 305–331.

[7]. BOHNENBLUST, H. & KARLIN, S., Geometrical properties of the unit sphere of Banach algebras. *Ann. Math.*, 62 (1955), 217–229.

[8]. BONSALL, F. F., On the representation of the points of a convex set. *J. London Math. Soc.* (A paraître.)

[9]. CHOQUET, G., Existence et unicité des représentations intégrales. *Séminaire Bourbaki* (Décembre, 1956), 139, 15 pages.

[10]. —— Le theorème de représentation intégrale dans les ensembles convexes compacts. *Ann. Inst. Fourier, Grenoble*, 10 (1960), 333–344.

[11]. —— Ensembles et cones convexes faiblement complets. I et II. *C. R. Acad. Sci. Paris*, 254 (1962), 1908–1910 et 2123–2125.

[12.] —— Remarques à propos de la démonstration d'unicité de P. A. Meyer. *Séminaire du Potentiel, Paris*, 6 (1962), n° 8, 13 pages.

[13]. —— Axiomatique des mesures maximales. *C.R. Acad. Sci. Paris*, 255 (1962), 37–39

[14]. —— Etude des mesures coniques. Cônes convexes saillants faiblement complets sans génératrices extrémales, *C.R. Acad. Sci. Paris*, 255 (1962), 445–447.

[15]. —— Frontière associée à un espace vectoriel de fonctions continues. I et II. *C.R. Acad. Sci. Paris*, 255, (1962), 816–818 et 1055–1057.

[16]. CHOQUET, G. & MEYER, P. A., Existence et unicité des représentations intégrales dans les convexes compacts quelconques. *Ann. Inst. Fourier, Grenoble*, 13 (1963),

[17]. GEL'FAND, I. M. & VILENKIN, N. I., *Fonctions généralisées*. IV, Ch. IV. Moscou, 1961.

[18]. GLICKSBERG, I., The representation of functionals by integrals. *Duke Math. J.*, 19 (1952), 253–261.

[19]. HERVÉ, M., Sur les représentations intégrales à l'aide des points extrémaux. *C.R. Acad. Sci. Paris*, 253 (1961), 366–368.

[20]. KADISON, R., A representation theory for commutative topological algebra. *Mem. Amer. Math. Soc.*, no. 7.

[21]. KLEE, V. L., Extremal structure of convex sets. II. *Math. Z.*, 69 (1958), 90–104.

[22]. MEYER, P. A., Sur les démonstrations nouvelles du théorème de Choquet. *Séminaire du Potentiel, Paris*, 6 (1962), n° 7, 9 pages. Voir aussi [16].

[23]. MILMAN, D. P., *Dokl. Akad. Nauk S.S.S.R., nouv. série*, 57 (1947), 119–122.

[24]. MOKOBODZKI, G., Balayage défini par un cône convexe de fonctions. *C.R. Acad. Sci. Paris*, 254 (1962), 803–805.

[25]. MOKOBODZKI, G., Principe du balayage, principe de domination. *Séminaire d'Initiation à l'Analyse*, 1ière Année (1962), n° 1, 11 pages.

[26]. WERNER, J., On algebras of continuous functions. *Proc. Amer. Math. Soc.*, 4 (1953), 866–869.

IDEMPOTENT MEASURES AND HOMOMORPHISMS OF GROUP ALGEBRAS

By PAUL J. COHEN

1. Introduction

In the last few years, attention has been given to problems in harmonic analysis which give rise to difficult questions in classical analysis, yet which are motivated most naturally from an abstract point of view. Indeed some of these problems had been posed by classical analysts, although their solutions have, even if not explicity, appealed to more modern ideas. I shall concern myself in this talk primarily with the problems referred to in the title but I should like to begin by discussing some other questions.

The more algebraic point of view may be said to have properly begun with the proof by Wiener [23] of the Generalized Tauberian theorem. This, as is well known, states that a closed subspace I of L^1 of the real line, closed under translation, either is all of L^1 or else there is a common zero of the Fourier transforms of the functions in I. The method of proof used by Wiener is completely classical although he uses ideas which in retrospect seem to anticipate the abstract approach. One can reformulate Wiener's theorem by saying that every proper closed ideal of $L^1(R)$ (with convolution as the ring multiplication) is contained in a regular maximal ideal.

Beurling [2], treated the problem of extending this result in the form now known as "spectral synthesis" and was able to solve it in some weaker forms.

Simply stated the problem of spectral synthesis asks "What are the closed ideals of $L^1(G)$?" More precisely, does every closed ideal in $L^1(G)$ consist of all f such that $\hat{f}=0$ on some closed set E of $\hat{G}$? (Here $\hat{f}$ is the Fourier transform of f and $\hat{G}$ is the dual group to G.) Equivalently, is every closed ideal the intersection of regular maximal ideals? One sees that in this form it would seem a very natural question from an algebraic point of view, although as I have indicated, the solution is completely analytical in nature and involves difficult questions in classical analysis.

The problem remained unsolved, except in some special cases, until Schwartz [21] gave a remarkably simple example to show that in general in R^3 spectral synthesis was impossible. Despite its simplicity this example showed the great power of the theory of distributions in suggesting new points of view. More recently, Malliavin [17] has disproved spectral synthesis for general locally compact Abelian groups. The difficulties encountered in the construction of an example for R^1, say, are very far from being merely technical and he used in a very beautiful way both the idea of distributions and a technique which in a different form had been used by Kahane [14] and Katznelson [15].

A second problem also has its roots in the Wiener Tauberian theorem. Since $L^1(R)$ is a ring it of course follows that if $p(z)$ is a polynomial and $f \in L^1(R)$ there is an obvious meaning to the statement $g=p(f)$. This means that $\hat{g}=p(\hat{f})$. In the course of his proof Wiener essentially proved that if

$p(z)$ is any analytic function whose domain of definition includes the support of $\hat{f}$, then there exists $g \in L^1(R)$ such that $\hat{g} = p(\hat{f})$. The essential difficulty here is proving that in the ring of absolutely convergent Fourier series the function $1/z$ has this property. This theorem is usually referred to as the Wiener-Lévy theorem, the final form first being stated by Lévy [16]. More recently the theorem was generalized by Šilov to functions p of several variables [21], an inessential restriction in his formulation being removed by Arens and Calderon [1]. The problem of which functions $\varphi(z)$ are such that if $f \in L^1$, there exists $g \in L^1$ such that $\hat{g} = \varphi(\hat{f})$ first drew attention when Littlewood suggested the problem of finding a Fourier cosine series $\sum a_n \cos nx$ such that $\sum a_n \sin nx$ is not a Fourier series. This problem corresponds to the problem of determining whether $\varphi(z) = |z|$ operates on L^1 of the circle group as above. Kahane [13] solved this problem and the general problem was first solved by Katznelson [15] who supplied the missing link in Kahane's argument. The more general questions of functions operating on Fourier-Stieltjes transforms soon was answered and an excellent exposition of all these questions may be found in [11].

2. Idempotent measures

The third problem, which we shall treat in more detail, is the problem of idempotent measures. We recall that if G is a locally compact Abelian group, the finite measures form a normed ring where $\|\mu\|$ is the total variation of μ and $\mu * \nu$ is defined by

$$\int f\, d(\mu * \nu) = \int\int f(x+y)\, d\mu(x) d\nu(y),$$

for all continuous f with compact support. One may ask, which μ satisfy $\mu * \mu = \mu$? If one passes to Fourier-Stieltjes tranforms, we have $\hat{\mu} = 0$ or 1.

The first successful attack on this problem was made by Helson [8, 10], who showed that the only Fourier series with coefficients either zero or one are of the form $\sum a_n e^{inx}$ where for some k, we have $a_n = a_{n+k}$ for all but finitely many n. His method uses ideas in complex-function theory, in particular the celebrated F. and M. Riesz theorem, (which has now been generalized by Helson and Lawdenslager [12]), and did not generalize to several variables. Using the results of [12], Rudin [20] did generalize the theorem to several variables. One may remark that an interesting by-product of Helson's work was the solution of a conjecture of Steinhaus on trigonometrical series whose partial sums are all positive [9].

To state the general theorem we first note that for any finite N, $(\sum_{j=1}^{N} e^{in_j x})\, dx$ is an idempotent measure on the circle. It is also simple to see that if H is a closed subgroup of G, then any idempotent measure on H considered as a measure on G is also idempotent. Hence

Definition. A measure μ on a group G is called a primitive idempotent measure, if for some compact subgroup H, we have

$$\mu = (\chi_1 + \ldots + \chi_r) dm,$$

where χ_i are characters of H and dm is normalized Haar measure on H.

The set of all idempotent measures, as do idempotents in any commutative ring, form a Boolean ring under the operations of

$$\mu \cup \nu = \mu + \nu - \mu * \nu \quad \text{and} \quad \mu \cap \nu = \mu * \nu.$$

We know that $\hat{\mu}$ must be the characteristic function of a set E and one sees that these operations correspond to union and intersection of the sets E. The main theorem is:

THEOREM 4. *The set of idempotent measures on G is precisely the Boolean ring generated by the primitive idempotent measures.*

The main difficulty in the course of the proof is the analysis of the singular part of the measure, since it is easily seen that if μ is absolutely continuous and idempotent then it is primitive. One sees easily first that,

THEOREM. *If μ is idempotent, then the support of μ lies in some compact subgroup H.*

Proof. Since $\hat{\mu}$ is either zero or one and is uniformly continuous, for some open neighborhood V of 0 in $\hat{G}$, we must have $\hat{\mu}(x+y) \equiv \hat{\mu}(x)$ if $y \in V$. This means that $\hat{\mu}$ is invariant under translation by the open subgroup generated by V, or by duality that μ has support in the compact subgroup of G which is its annihilator.

After some manipulation which is of a fairly formal character one can show that the theorem will follow if the following lemma is proved:

LEMMA. *Let μ be a measure in a compact subgroup such that $\hat{\mu}$ takes only integral values. Then either μ is absolutely continuous, or for some closed subgroup H of infinite index, we have $|\mu|(H) \neq 0$.*

The proof of this lemma requires a fairly involved argument of a combinational nature and rather than give much detail, we prefer to give an example of the type of reasoning involved. We show that

THEOREM. *If μ is idempotent and $\|\mu\| = 1$, then $\mu = \chi\, dm$ where χ is a character and dm is Haar measure of a closed subgroup.*

Proof. Let S be the subset of $\hat{G}$ where $\hat{\mu} = 1$. We show that if χ_1, χ_2 and $\chi_3 \in S$ then $\chi_1 \chi_2 \chi_3^{-1} \in S$.

Consider the function

$$f = \tfrac{1}{3}(\chi_1 + \chi_2 + \chi_3) - c(\chi_1 \chi_2 \chi_3^{-1}) \quad \text{where } c > 0. \tag{1}$$

It is clear that if c is chosen small enough, we have $|f| \leqslant 1 - \delta$ for some δ. Then we have $|\int f d\mu| \leqslant 1 - \delta$. But, on the other hand since $\hat{\mu} = 1$ on S and $\hat{\mu} = 0$ off S we have

$$\int d\mu = \begin{cases} 1 - c & \text{if} \quad \chi_1 \chi_2 \chi_3^{-1} \in S, \\ 1 & \text{if} \quad \chi_2 \chi_2 \chi_3^{-1} \notin S. \end{cases}$$

Hence $\chi_1 \chi_2 \chi_3^{-1} \in S$, and so S is a coset of some subgroup K of $\hat{G}$ and the theorem easily follows. Rudin has computed the best constants in this argument and has shown that the conclusion holds if $\|\mu\| \leqslant \sqrt{5/2}$.

In the proof of the general results one gives a systematic way to construct more complicated expressions than (1) which are then used to show that the set S satisfies certain combinatorial conditions.

In [4], it is shown how there ideas may be used to attack a problem of Littlewood. We have seen that in the case of Fourier series, the simplest idempotents are the finite exponential sums. The next simplest example is Haar measure on a proper subgroup of the circle, whose Fourier-Stieltjes series must be of the form,

$$\sum_{n=\infty}^{\infty} e^{iknx}$$

for some integer k. The norm of such a measure is of course 1. This would seem to suggest that the smallest norms of exponential sums are encountered among sums of exponentials in an arithmetic progression. This bears some relation to results of Hardy and Littlewood concerning rearrangements of Fourier series. One may conjecture that since

$$\int \left| \sum_{1}^{N} e^{inx} \right| dx \sim A \log N,$$

one also has

$$\int \left| \sum_{j=1}^{N} e^{in_j x} \right| dx > B \log N. \tag{2}$$

This conjecture remains unproved, but it is shown in [4], that

$$\int \left| \sum_{j=1}^{N} e^{in_j x} \right| dx > B \left(\frac{\log N}{\log \log N} \right)^{\frac{1}{8}}. \tag{3}$$

Davenport [6], by modifying one of the combinatorial lemmas, has shown that the exponent in the right side of (3) may be replaced by 1/4. The inequality (3) may be readily generalized to sums of characters on arbitrary connected Abelian groups.

3. Homomorphisms of group algebras

The problem of determining the homomorphisms of $L^1(G) \to L^1(H)$ was treated by Beurling and Helson in [3] for the case $G = H = R$, in a slightly different form. They were able to prove that every such homomorphism corresponds in a natural way to a homomorphism $x \to \lambda x$ of R into R.

It is not difficult to show, by means of the closed graph theorem, that any homomorphism $\varphi: L^1(G) \to L^1(H)$ is continuous and is determined by its adjoint map $\varphi^*: \hat{H} \to \{\hat{G}, 0\}$ where 0 represents the "trivial character" mapping G identically onto zero. For example, the theorem in the case of the real line reduces to proving

THEOREM. *Let $\omega(x)$ be a function such that whenever $\hat{f}(x)$ is a Fourier transform, then $\hat{f}(\omega(x))$ is a Fourier transform. Then $\omega(x) = ax + b$.*

The same statement holds for absolutely convergent Fourier series. The case when $\hat{H}$ is a disconnected group is much more interesting. For H the

circle group this was solved by Rudin [18, 19]. His method used the results concerning idempotent measures in the circle. However, it was not noticed that there is a simple relation which more or less reduces the problem of homormorphisms to that of idempotent measures. Namely,

THEOREM 5. *Let $\varphi: L^1(G) \to L^1(H)$ be a homomorphism, $\varphi_*: \hat{H} \to (\hat{G}, 0)$ the dual map. If G and H are compact, and S the subset of $\hat{G} \times \hat{H}$ consisting of all pairs $(\chi, \varphi_*(\chi))$ where $\varphi_*(\chi) \neq 0$, then the characteristic function of S is the Fourier transform of an idempotent measure.*

The case of non-compact G and H must be handled by passing to their Bohr compactifications. The final result is stated in terms of affine maps.

Definition. Given a continuous map α from a coset K of a subgroup G' of G into a group H, we say α is affine if $\alpha(x+y-z) = \alpha(x) + \alpha(y) - \alpha(z)$ for x, y and z in K.

THEOREM. *The map φ_* is the adjoint map of a homomorphism $\varphi: L^1(G) \to M(H)$, where $M(H)$ are the measures on H, if and only if there is a decomposition of $\hat{H}$ into sets S_j, belonging to the Boolean ring generated by a finite number of cosets K_i of open subgroups of $\hat{H}$ with the property that for each S_j, there exists a K_i with $K_i \supset S_j$ and φ_* on S_j is either identically zero, or agrees with an affine map of K_i.*

The map φ takes $L^1(G)$ into $L^1(H)$ if and only if φ_*^{-1} of every compact subset of $\hat{G}$ is compact.

It is not hard to see that this includes the corresponding results for the special cases of the real line and the circle. Using these results one can prove for example that the subalgebra $L^1(G)$ is mapped onto itself by any automorphism of $M(G)$. Another illustration is the theorem of Beurling-Helson [3] which says that if f is a Fourier transform of a measure on the line such that f^n for positive and negative n are the Fourier transforms of uniformly bounded measures then $f(x) = e^{i(ax+b)}$. One can also obtain results which give conditions under which one may deduce from $L^1(G)$ being isomorphic to $L^1(H)$ that G and H are isomorphic. In particular this holds if $\hat{H}$ is connected.

In conclusion, it should be pointed out that almost no results of this type are known for the case of non-Abelian groups. It would seem that compact Abelian groups might offer some hope for suitable generalizations.

REFERENCES

[1]. ARENS, R. & CALDERON, A. P., Analytic functions of several Banach algebra elements. *Ann. Math.*, 62 (1955), 204–216.

[2]. BEURLING, A., On the spectral synthesis of bounded functions. *Acta Math.*, 81 (1949) 225–238.

[3]. BEURLING, A. & HELSON, H., Fourier-Stieltjes transforms with bounded powers. *Math. Scand.*, 1 (1953) 120–126.

[4]. COHEN, P. J., On a conjecture of Littlewood and idempotent measures. *Amer. J. Math.*, 82 (1960), 191–212.

[5]. —— On homomorphisms of group algebras. *Amer. J. Math.*, 82 (1960), 213–226.

[6]. DAVENPORT, H., On a theorem of P. J. Cohen. *Mathematika* 7 (1960), 93–97.

[7]. HELSON, H., Isomorphisms of abelian group algebras. *Ark. Mat.*, 2 (1953), 475–487.

[8]. —— Note on harmonic functions. *Proc. Amer. Math. Soc.*, 4 (1953), 686–691.

[9]. —— Proof of a conjecture of Steinhaus. *Proc. Nat. Acad. Sci. U.S.A.*, 40 (1954), 205–206.

[10]. —— On a theorem of Szegö. *Proc. Amer. Math. Soc.*, 6 (1955), 235–242.

[11]. HELSON, H., KAHANE, J. P., KATZNELSON, Y. & RUDIN, W., The functions which operate on Fourier transforms. *Acta. Math.*, 102 (1959), 135–157.

[12]. HELSON, H. & LAWDENSLAGER, D., Prediction theory and Fourier series in several variables. *Acta Math.*, 99 (1958), 165–202.

[13]. KAHANE, J. P., Sur un problème de Littlewood. *Nederl. Akad. Wetensch. Indag. Math.*, 19 (1957), 268–271.

[14]. —— Sur un théorème de Wiener-Lévy. *C.R. Acad. Sci. Paris*, 246 (1958), 1949–1951.

[15]. KATZNELSON, Y., Sur les fonctions opérant sur l'algèbre des séries de Fourier absolument convergentes. *C.R. Acad. Sci. Paris*, 247 (1958), 404–406.

[16]. LÉVY, P., Sur la convergence absolue des séries de Fourier. *Compositio Math.*, 1 (1934), 1–14.

[17]. MALLIAVIN, P., Sur l'impossibilité de la synthèse spectrale sur la droite. *C.R. Acad. Sci. Paris*, 248 (1959), 2155–2157.

[18]. RUDIN, W., The automorphisms and endomorphisms of the group algebra of the unit circle. *Acta Math.*, 95 (1956), 39–56.

[19]. —— On isomorphisms of group algebras. *Bull. Amer. Math. Soc.*, 64, (1958), 167–169.

[20]. —— Idempotent measures on abelian groups. *Pracific J. Math.*, 9 (1959), 195–209.

[21]. SCHWARTZ, L., Sur une propriété de synthèse spectrale dans les groupes noncompacts. *C.R. Acad. Sci. Paris*, 227 (1948), 424–426.

[22]. ŠILOV, G. E., On the decomposition of a commutative normed ring as a direct sum of ideals. *Mat. Sb. N.S.*, 32 (1953), 353–364.

[23]. WIENER, N., Tauberian theorems. *Ann. Math.*, 33 (1932), 1–100.

CONDITIONALLY CONVERGENT FUNCTIONAL INTEGRALS AND PARTIAL DIFFERENTIAL EQUATIONS

By LEON EHRENPREIS

The theory of functional integration had its beginning with the work of Einstein, Smolouchowsky, Paley, and Wiener on Brownian motion. Later Feyneman introduced a process which bears the same relation to the Schrödinger equation that Brownian motion bears to the heat equation. Feyneman's method has never been subjected to a rigorous mathematical treatment. In the present lecture we shall describe a procedure for solving the initial value problem and the Cauchy problem for a wide class of partial differential equations. This method may be thought of as a general type of functional integration. In the special cases of the heat and Schrödinger equations we obtain the above-mentioned techniques of Einstein *et al.* and of Feyneman, respectively. Our method is mathematically rigorous and can therefore be used to establish existence theorems for initial value and Cauchy problems. For example, we can give a new proof of the Cauchy–Kowalewski theorem (actually in a somewhat stronger form as first derived by Pucci). We give also a new proof of the Leray–Petrowski existence theorem for linear hyperbolic equations. This proof extends to some cases with multiplie characteristics.

The main idea of our method of functional integration is to obtain the solution of the initial value or Cauchy problem for a system with variable coefficients as a limit of functions obtained by solving constant coefficient problems.

Formally we proceed as follows: Let $x=(x_1,\ldots,x_n)$, and consider the initial value problem $\partial f/\partial t = A(t,x,\partial/\partial t,\partial/\partial x)f$ with $f(0,x)=f^0(x)$. Here $f=f(t,x)$ is a vector of functions $f_1,\ldots,f_m$, and A is an $m\times m$ matrix, and $f^0(x)$ is a given vector of functions. Suppose we want to find $f(1,x)$. We break up the t interval $[0,1]$ into N equal parts. Then we define the approximate solution $f^N(t,x)$ as follows: for any t_0 in $[0,1/N]$ and any x_0 we solve the initial value problem

$$\partial f^N_{t_0,x_0}(t,x)/\partial t = A(t_0,x_0,\partial/\partial t,\partial/\partial x)\, f^N_{t_0,x_0}(t,x)$$

with initial value $f^N_{t_0,x_0}(0,x)=f^0(x)$. We thus obtain a family of functions $f^N_{t_0,x_0}(t,x)$. We define $f^N(t_0,x_0)=f^N_{t_0,x_0}(t_0,x_0)$ for $0\leqslant t_0\leqslant 1/N$.

Next for $1/N<t_0\leqslant 2/N$ and any x_0 we consider the initial value problem

$$\partial f^N_{t_0,x_0}(t,x)/\partial t = A(t_0,x_0,\partial/\partial t,\partial/\partial x)\, f^N_{t_0,x_0}(t,x)$$

with initial value $f^N_{t_0,x_0}(1/N,x)=f^N(1/N,x)$. We then define

$$f^N(t_0,x_0)=f^N_{t_0,x_0}(t_0,x_0) \text{ for } 1/N<t_0\leqslant 2/N.$$

Proceeding in this way we construct $f^N(t,x)$ for $0 \leqslant t \leqslant 1$.

Finally we must find conditions on A and suitable topologies so that f^N converges to a solution of the given initial value problem. In most cases the procedure must be modified slightly in order to obtain convergence.

By using explicit formulae for the solution of the initial value problems for the partial differential equations with constant coefficients for the $f^N_{t_0, x_0}$ we can interpret the limit of the f^N as a functional integral.

EXISTENCE, UNIQUENESS AND REGULARITY OF SOLUTIONS OF LINEAR DIFFERENTIAL EQUATIONS

By LARS HÖRMANDER

The aim of this lecture is to report some recent results concerning existence of solutions of linear differential equations of general type. The main emphasis will be placed on theorems concerning uniqueness and regularity of the solutions of a differential equation, for such results imply existence theorems for the adjoint equation (see section 3). We shall also exhibit classes of differential equations which do not possess solutions.

1. Notations

Let Ω be a C^∞ manifold of dimension n which is countable at infinity, and let P be a linear differential operator on Ω with C^∞ coefficients. This means (Peetre [10]) that P is a linear operator in $C^\infty(\Omega)$ which contracts supports,

$$\operatorname{supp} Pf \subset \operatorname{supp} f \quad (f \in C^\infty(\Omega)). \tag{1.1}$$

By continuity P can be extended to a linear operator, also denoted by P, in the space $\mathcal{D}'(\Omega)$ of distributions on Ω.

Let $d\Omega$ be a fixed positive C^∞ density in Ω. The identity (1)

$$\int (Pu)\,\bar{v}\,d\Omega = \int u\overline{P^* v}\,d\Omega \text{ when } u,\ v \in C^\infty(\Omega) \text{ and } \operatorname{supp} u \cap \operatorname{supp} v \Subset \Omega, \tag{1.2}$$

then defines a differential operator P^*, called the formal adjoint of P (with respect to the density $d\Omega$).

The operator P is of order $\leqslant m$ if and only if the polynomial

$$\tau \to e^{-i\tau\varphi} P\, e^{i\tau\varphi}$$

is of degree $\leqslant m$ for every $\varphi \in C^\infty(\Omega)$. The coefficient of τ^m is then for every $x \in \Omega$ a form of degree m in grad $\varphi(x)$, which we denote by $P_m(x, \operatorname{grad} \varphi)$; if the order of P is equal to m it is called the characteristic form of P.

If $A \subset \Omega$ we write

$$C_0^\infty(A) = \{u;\ u \in C^\infty(\Omega),\ \operatorname{supp} u \Subset A\},$$

and $\mathcal{E}'(A)$ is defined similarly with $C^\infty(\Omega)$ replaced by $\mathcal{D}'(\Omega)$. Finally, if $u \in \mathcal{D}'(\Omega)$ we denote by sing supp u the smallest closed subset of Ω such that u is a C^∞ function in its complement.

(1) The notation $A \Subset B$ means that A is relatively compact in B.

2. Equations without solutions

Lewy [6] discovered that the differential equation

$$\partial u/\partial x_1 + i\,\partial u/\partial x_2 + 2i(x_1 + ix_2)\partial u/\partial x_3 = f$$

does not have a solution in any open subset of R_3 for a suitable choice of $f \in C^\infty(R_3)$. This result can be generalized as follows (see Hörmander [3, Chap. VI]):

THEOREM 2.1. *Let P be of order $\leqslant m$, which implies that the commutator $C = P^*P - PP^*$ is of order $\leqslant 2m-1$. Assume that for some $x \in \Omega$ and some real covariant vector ξ at x we have*

$$P_m(x,\xi) = 0, \quad C_{2m-1}(x,\xi) \neq 0. \tag{2.1}$$

Then one can choose $f \in C^\infty(\Omega)$ such that the equation $Pu = f$ does not have a distribution solution in any neighborhood of x. If for every x in a dense subset of Ω the hypothesis (2.1) is fulfilled for some real ξ, one can find $f \in C^\infty(\Omega)$ such that the equation $Pu = f$ does not have a distribution solution in any open subset of Ω.

The main part of the proof consists in constructing functions which satisfy the equation $P^*v = 0$ except for a small error and which are strongly concentrated at the point x. This is done by means of asymptotic expansions closely related to those of diffraction theory.

3. Existence theorems

Results on regularity and uniqueness for the adjoint P^* of P imply existence theorems for P. Indeed, we have the following two theorems.

THEOREM 3.1. *Assume that*

$$v \in \mathcal{E}'(\Omega),\ P^*v \in C_0^\infty(\Omega) \Rightarrow v \in C_0^\infty(\Omega). \tag{A}$$

Then the vector space

$$N(K) = \{v;\ v \in C_0^\infty(K),\ P^*v = 0\}$$

is finite dimensional for every compact set $K \subset \Omega$. If $f \in \mathcal{D}'(\Omega)$ and

$$\tilde{f}(v) = 0 \quad (v \in N(K)), \tag{3.1}$$

there exists a distribution $u \in \mathcal{D}'(\Omega)$ such that $Pu = f$ in the interior of K.(1)

To prove this theorem one first uses the closed graph theorem to give a quantitative statement of the hypothesis (A). Then it follows immediately that $N(K)$ is finite dimensional and that P^* defines an isomorphism of $C_0^\infty(K)/N(K)$ into $C_0^\infty(K)$. By (3.1) the linear form

$$P^*v \to \tilde{f}(v) \quad (v \in C_0^\infty(K)),$$

is therefore continuous for the topology in $C^\infty(\Omega)$. Hence it can be extended to a distribution $\bar{u} \in \mathcal{E}'(\Omega)$, and since $\bar{u}(P^*v) = \tilde{f}(v)$, $v \in C_0^\infty(K)$, we have $Pu = f$ in the interior of K.

(1) The author owes this result to F. Trèves.

THEOREM 3.2. *Assume in addition to the condition* (A) *that to every compact set* $K \subset \Omega$ *there is another compact set* $K' \subset \Omega$ *such that*

$$v \in \mathcal{E}'(\Omega), \text{ supp } P^*v \subset K \Rightarrow \text{supp } v \subset K', \tag{B}$$

$$v \in \mathcal{E}'(\Omega), \text{ sing supp } P^*v \subset K \Rightarrow \text{sing supp } v \subset K'. \tag{C}$$

Then the vector space

$$N = \{v;\, v \in C_0^\infty(\Omega),\ P^*v = 0\}$$

is finite dimensional and the equation $Pu = f$ *has a solution* $u \in \mathcal{D}'(\Omega)$ *for every* $f \in \mathcal{D}'(\Omega)$ *such that*

$$\tilde{f}(v) = 0 \quad (v \in N). \tag{3.2}$$

The Hahn-Banach theorem reduces the proof to showing that the linear form

$$P^*v \to \tilde{f}(v) \quad (v \in C_0^\infty(\Omega)),$$

is continuous for the Schwartz topology in $C_0^\infty(\Omega)$ when (3.2) is valid so that the form is uniquely defined. The continuity can be established by a slight modification of the proof of Theorem 3.6.4 in [3].

Simple examples show that the condition (A) is not necessary for the conclusion of Theorem 3.1 to hold; a weaker sufficient condition is that for some integer k we have $v \in C_0^\infty(\Omega)$ if $P^*v \in C_0^\infty(\Omega)$ and $v \in C_0^k(\Omega)$. However, (A) is fulfilled in most existence theorems known. Condition (B) in Theorem 3.2 is not necessary either, but if $P\mathcal{D}'(\Omega) \supset C^\infty(\Omega)$ then (B) must hold at least when $v \in C_0^\infty(\Omega)$. When $\Omega \subset R_n$ and the coefficients of P are constant, it follows that the full condition (B) is necessary in order that $P\mathcal{D}'(\Omega) \supset C^\infty(\Omega)$; conversely (B) then implies the existence of solutions of the equation $Pu = f$ when f is a distribution of finite order (Malgrange [7]). When P has constant coefficients and $\Omega \subset R_n$, condition (C) is also necessary in order that $P\mathcal{D}'(\Omega) = \mathcal{D}'(\Omega)$ (Hörmander [3]).

If P is a differential operator with constant coefficients in R_n, we have

$$H(\text{supp } v) = H(\text{supp } P^*v), \quad H(\text{sing supp } v) = H(\text{sing supp } P^*v) \quad (v \in \mathcal{E}'), \tag{3.3}$$

where H denotes convex hulls. The first identity is a special case of the theorem of supports, and both have simple proofs via Laplace transforms. (See John [5], Malgrange [8].) Hence $P\mathcal{D}'(\Omega) = \mathcal{D}'(\Omega)$ if Ω is convex. This result is due to Ehrenpreis [2] for $\Omega = R_n$ and to Malgrange [8] for arbitrary convex Ω.

In the following paragraphs we shall give *local* regularity and uniqueness theorems which lead to sufficient conditions for (A), (B) or (C) to be valid. (One can also give local statements which are equivalent to (3.3); see John [5], Malgrange [8].) The existence theorems thus obtained will not be stated explicitly, but they go beyond those mentioned above even for differential operators with constant coefficients.

4. Hypoelliptic operators

The operator P^* is called hypoelliptic if

$$\text{sing supp } v = \text{sing supp } P^*v, \quad v \in \mathcal{D}'(\Omega). \tag{D}$$

This local property of P^* trivially implies both (A) and (C). It is classical (Petrowsky [11]) that elliptic operators are hypoelliptic, and so are diffusion operators. Algebraic conditions characterizing hypoelliptic differential operators with constant coefficients in open sets $\Omega \subset R_n$ are known (see Hörmander [3]). The hypoellipticity of a related class of differential operators with variable coefficients, including elliptic operators, has been proved by Hörmander and by Malgrange (see [3]). Weaker sufficient conditions for hypoellipticity have been found by Trèves [13] and by Hörmander [4], but so far no sufficient algebraic condition for hypoellipticity has been found which is invariant for coordinate transformations and is weaker than ellipticity. (See Trèves [14].)

It should be noted that condition (B) is not always fulfilled even for elliptic operators. Indeed, Pliš [12] has constructed an elliptic differential equation which has a solution with compact support. This leads easily to examples of elliptic operators which do not satisfy (B).

5. Holmgren's uniqueness theorem

From now on we always assume that $\Omega \subset R_n$, but since the results we shall give are invariant for changes of variables, they are also applicable to manifolds Ω. We assume that P has a finite order which we denote by m Then the classical uniqueness theorem of Holmgren is as follows.

THEOREM 5.1. *Let φ be a real valued function in $C^1(\Omega)$ and let x^0 be a point in Ω where*

$$P_m(x^0, \operatorname{grad} \varphi(x^0)) \neq 0, \tag{5.1}$$

that is, the surface $\varphi(x) = \varphi(x^0)$ is non-characteristic. If the coefficients of P are real analytic, there exists a neighborhood $\Omega' \subset \Omega$ of x^0 such that every $u \in \mathcal{D}'(\Omega)$ satisfying the equation $Pu = 0$ and vanishing in $\Omega^+ = \{x;\ x \in \Omega, \varphi(x) > \varphi(x^0)\}$ must also vanish in Ω'.

For a proof valid for distribution solutions we refer to Hörmander [3], where the following supplementary result is also given.

THEOREM 5.2. *Let φ be a real valued function in $C^2(\Omega)$ and let $P(x, D)$ be a differential operator with analytic coefficients and with real coefficients in the principal part. Let x^0 be a point in Ω where* grad $\varphi(x^0) = N^0 \neq 0$ *and $P_m(x^0, N^0) = 0$ but*

$$\sum_{j,k=1}^{n} \frac{\partial^2 \varphi(x^0)}{\partial x_j \partial x_k} P_m^{(j)}(x^0, N^0) P_m^{(k)}(x^0, N^0) + \sum_{1}^{n} P_m^{(k)}(x^0, N^0) P_{m,k}(x^0, N^0) > 0. \tag{5.2}$$

Here P_m denotes the principal part of P and we have used the notation

$$P_m^{(j)}(x,\xi) = \partial P_m(x,\xi)/\partial \xi_j, \quad P_{m,j}(x,\xi) = \partial P_m(x,\xi)/\partial x_j. \tag{5.3}$$

Then we obtain the same conclusion as in Theorem 5.1.

On the other hand, Malgrange [9] has shown that there is non-uniqueness if the inequality opposite to (5.2) holds. The meaning of (5.2) is perhaps more clear if we note that the left-hand side is the second derivative of the restriction of φ to the bicharacteristic with initial data $x = x^0$, $\xi = N^0$.

6. Carleman estimates

In this paragraph we shall discuss a type of estimates first studied by Carleman in connection with uniqueness theorems for equations with non-analytic coefficients. These estimates are of the form

$$\tau\int|D^{m-1}u|^2e^{2\tau\varphi}\,dx\leqslant K_1\int|Pu|^2e^{2\tau\varphi}\,dx$$
$$+K_2\sum_0^{m-2}\tau^{2(m-j)-1}\int|D^ju|^2e^{2\tau\varphi}\,dx,\quad (u\in C_0^\infty(\Omega)). \tag{6.1}$$

Here τ is a large positive parameter, φ is a fixed weight function in $C^2(\Omega)$ and $|D^ju|^2$ is the sum of the squares of the absolute values of the derivatives of u of order $\leqslant j$. Such estimates with $K_2=0$ have a particular interest as we shall see in section 7. For proofs of all the following results we refer to Hörmander [3, Chap. VIII].

THEOREM 6.1. *Let* $N=\operatorname{grad}\varphi(x)$ *where* $x\in\Omega$, *and let* $\zeta=\xi+i\sigma N$ *with* $\xi\in R_n$ *and* $0\neq\sigma\in R_1$ *satisfy the characteristic equation*

$$P_m(x,\zeta)=0. \tag{6.2}$$

If (6.1) *is valid, it follows that*

$$|\zeta|^{2(m-1)}-K_2\sigma^2(|\zeta|^2+\sigma^2)^{m-2}\leqslant 2K_1\varphi_P''(x,\xi,\sigma) \tag{6.3}$$

when the left-hand side is positive. Here we have used the notation

$$\varphi_P''(x,\xi,\sigma)=\sum_{j,k=1}^n\frac{\partial^2\varphi}{\partial x_j\partial x_k}P_m^{(j)}(x,\zeta)\,\overline{P_m^{(k)}(x,\zeta)}$$
$$+\sigma^{-1}\operatorname{Im}\sum_1^n P_{m,k}(x,\zeta)\,\overline{P_m^{(k)}(x,\zeta)}. \tag{6.4}$$

Note that if K_2 is a large positive constant we only obtain a condition on $\varphi_P''(x,\xi,\sigma)$ when $\zeta=\xi+i\,\sigma\,N$ is nearly real.

The proof of Theorem 6.1 is achieved by applying (6.1) to functions of the form

$$u(y)=\exp(i\,\tau\,w(y)/\sigma)\psi((y-x)\sqrt{\tau})$$

where $\psi\in C_0^\infty(R_n)$, $w\in C^\infty(R_n)$ and $w(y)=\langle y-x,\zeta\rangle+O(|y-x|^2)$ when $y\to x$. Letting $\tau\to\infty$ leads to an inequality, involving ψ and its first order derivatives, which is easily seen to be equivalent to (6.3).

When $\sigma\to 0$ it is easy to see that (6.3) implies

$$C_{2m-1}(x,\xi)=0\quad\text{if } x\in\Omega,\ \xi\in R_n\quad\text{and}\quad P_m(x,\xi)=0 \tag{6.5}$$

(cf. Theorem 2.1), for $\varphi_P''(x,\xi,\sigma)-C_{2m-1}(x,\xi)/2\sigma$ is a polynomial in ξ and σ. To proceed we use a strengthened form of (6.5), but it is quite possible that this additional assumption can be eliminated to a large extent.

Definition. P is principally normal if there is a differential operator Q of order $m-1$ with C^1 coefficients such that

$$C_{2m-1}(x,\xi)=2\operatorname{Re}P_m(x,\xi)\overline{Q_{m-1}(x,\xi)},\quad \xi\in R_n. \tag{6.6}$$

It is obvious that P is principally normal if P_m has real or constant coefficients. When $P_m(x,\xi+i\sigma N)=0$ we now modify (6.4) as follows

$$\varphi_P''(x,\xi,\sigma)=\sum_{j,k=1}^{n}\frac{\partial^2\varphi}{\partial x_j\partial x_k}P_m^{(j)}(x,\zeta)\overline{P_m^{(k)}(x,\zeta)}$$
$$+\sigma^{-1}\left\{\operatorname{Im}\sum_1^n P_{m,k}(x,\zeta)\overline{P_m^{(k)}(x,\zeta)}-\operatorname{Re}P_m(x,\zeta)\overline{Q_{m-1}(x,\zeta)}\right\}.$$

Since the right-hand side is a polynomial in ξ and σ by (6.6), it can be taken as definition of $\varphi_P''(x,\xi,\sigma)$ when $\sigma=0$ and $P_m(x,\xi)=0$. If $\xi\in R_n$ and

$$P_m(x,\xi)=\sum_1^n P_m^{(j)}(x,\xi)N_j=0, \tag{6.7}$$

we obtain in particular

$$\varphi_P''(x,\xi,0)=\sum_{j,k=1}^{n}\frac{\partial^2\varphi}{\partial x_j\partial x_k}P_m^{(j)}(x,\xi)\overline{P_m^{(k)}(x,\xi)}$$
$$+\operatorname{Re}\sum_{j,k=1}^{n}(P_{m,k}^{(j)}(x,\xi)\overline{P_m^{(k)}(x,\xi)}-P_{m,k}(x,\xi)\overline{P_m^{(jk)}(x,\xi)})N_j. \tag{6.8}$$

When P_m has real coefficients, a simple computation shows that the right-hand side is the second derivative of the restriction of φ to the bicharacteristic with initial data x,ξ. (The first derivative of φ along this curve vanishes by (6.7).)

We can now state a partial converse of Theorem 6.1.

THEOREM 6.2. *Let P be principally normal or elliptic, let* grad $\varphi(x)\neq 0$, *and let $\varphi_P''(x,\xi,0)>0$ if $x\in\Omega$, $0\neq\xi\in R_n$ and $P_m(x,\xi)=0$. If K is a compact subset of Ω, we have for some constant C*

$$\sum_0^{m-1}\tau^{2(m-j)-1}\int|D^ju|^2e^{2\tau\varphi}dx\leqslant C\int(|Pu|^2+\tau^{2m-1}|u|^2)e^{2\tau\varphi}dx\quad(u\in C_0^\infty(K)). \tag{6.9}$$

If in addition $\varphi_P''(x,\xi,\sigma)>0$ for all $x\in\Omega$, $\xi\in R_n$ and $\sigma\in R_1$ with $\sigma\neq 0$. satisfying the equation $P_m(x,\xi+i\sigma N)=0$, we obtain

$$\sum_0^{m-1}\tau^{2(m-j)-1}\int|D^ju|^2e^{2\tau\varphi}dx\leqslant C\int|Pu|^2e^{2\tau\varphi}dx\quad(u\in C_0^\infty(K)). \tag{6.10}$$

In proving these estimates one first makes an integration by parts in $\int|Pu|^2e^{2\tau\varphi}dx$ ("the energy integral method"), and the resulting expression is then studied by means of Fourier transforms.

7. Uniqueness of the Cauchy problem

From the estimate (6.10) one obtains by standard arguments:

THEOREM 7.1. *Let φ be a real valued function in $C^2(\Omega)$ and let P be a diffe-*

rential operator which is either principally normal or elliptic. Let x^0 be a point in Ω where grad $\varphi(x^0)=N^0\neq 0$, *and assume that*

$$\varphi_P''(x^0,\xi,\sigma)>0 \tag{7.1}$$

if $0\neq\zeta=\xi+i\sigma N^0$, *where* $\xi\in R_n$ *and* $\sigma\in R_1$, *and* ζ *is a solution of the equations*

$$P_m(x^0,\zeta)=0,\quad \sum_1^n P_m^{(j)}(x^0,\zeta)N_j^0=0. \tag{7.2}$$

Then there is a neighborhood Ω' of x^0 such that every $u\in\mathcal{D}'(\Omega)$ satisfying the equation $Pu=0$ in Ω and vanishing in $\Omega^+=\{x;\, x\in\Omega,\varphi(x)>\varphi(x^0)\}$, must also vanish in Ω'.

If $P_m(x^0,N^0)=0$, the condition (7.1) with $\xi=N^0$ and $\sigma=0$ reduces to (5.2), but there is a striking contrast between the strength of the hypotheses we have made in the case of analytic coefficients and in the case of C^∞ coefficients.

Examples due to P. Cohen [1] show that the assumption (7.1) is vital at least when $\sigma=0$. Pliš [12] has given examples of non-uniqueness even for elliptic equations, but the role of the condition (7.1) when $\sigma\neq 0$ is not yet completely clarified.

8. Unique continuation of singularities

In sections 5 and 7 we have stated results which lead to sufficient conditions for (B) to be valid. We shall now give another consequence of Theorem 6.2 which can be used to verify (A) or (C).

THEOREM 8.1. *Let φ be a real valued function in $C^2(\Omega)$ and let P be principally normal. Let x^0 be a point in Ω where* grad $\varphi(x^0)=N^0\neq 0$, *and assume that*

$$\varphi_P''(x^0,\xi,0)>0 \tag{8.1}$$

if $0\neq\xi\in R_n$ *and*

$$P_m(x^0,\xi)=0,\quad \sum_1^n P_m^{(j)}(x^0,\xi)N_j^0=0. \tag{8.2}$$

Then there is a neighborhood Ω' of x^0 such that every $u\in\mathcal{D}'(\Omega)$ for which $Pu\in C^\infty(\Omega)$ and $u\in C^\infty(\Omega^+)$ is in $C^\infty(\Omega')$.

Results of Zerner [15] show that the convexity assumption (8.1) is essentially necessary for the conclusion to be valid.

REFERENCES

[1]. COHEN, P., The non-uniqueness of the Cauchy problem. *O.N.R. Techn. Report 93.* Stanford, 1960.

[2]. EHRENPREIS, L., Solutions of some problems of division III. *Amer. J. Math.*, 78 (1956), 685–715.

[3]. HÖRMANDER, L., Linear partial differential operators. *Grundl. d. math. Wiss.*, 116. (To appear.)

[4]. —— Hypoelliptic differential operators. *Ann. Inst. Fourier Grenoble*, 11 (1961), 477–492.

[5]. JOHN, F., Continuous dependence on data for solutions of partial differential equations with a prescribed bound. *Comm. Pure Appl. Math.*, 13 (1960), 551–585.

[6]. LEWY, H., An example of a smooth linear partial differential equation without solution. *Ann. Math.*, (2) 66 (1957), 155–158.

[7]. MALGRANGE, B., Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution. *Ann. Inst. Fourier Grenoble*, 6 (1955–56), 271–355.

[8]. —— Sur la propagation de la régularité des solutions des équations à coefficients constants. *Bull. Math. Soc. Sci. Math. Phys. R.P. Roumanie*, 3 (53) (1959), 433–440.

[9]. —— Sur les ouverts convexes par rapport à un opérateur différentiel. *C. R. Acad. Sci. Paris*, 254 (1962), 614–615.

[10]. PEETRE, J., Rectification à l'article "Une caractérisation abstraite des opérateurs différentiels". *Math. Scand.*, 8 (1960), 116–120.

[11]. PETROWSKY, I. G., Sur l'analyticité des solutions des systèmes d'équations différentielles. *Mat. Sb.*, 5 (47) (1939), 3–68.

[12]. PLIŠ, A., A smooth linear elliptic differential equation without any solution in a sphere. *Comm. Pure Appl. Math.*, 14 (1961), 599–617.

[13]. TRÈVES, F., Opérateurs différentiels hypoelliptiques. *Ann. Inst. Fourier Grenoble*, 9 (1959), 1–73.

[14]. —— An invariant criterion of hypoellipticity. *Amer. J. Math.*, 83 (1961), 645–668.

[15]. ZERNER, M., Sur les singularités des distributions vérifiants une équation aux derivées partielles à coefficients analytiques. *Coll. Int. du CNRS.* Paris, Juin 1962.

ON NORMALIZATION IN THE GENERAL COEFFICIENT THEOREM

By JAMES A. JENKINS

Teichmüller enunciated the intuitive principle that the solution of a certain type of extremal problem for univalent functions is determined by a quadratic differential for which the following prescriptions hold. If the competing mappings are to have a certain fixed point the quadratic differential will have a simple pole there. If in addition fixed values are required for the first n derivatives of competing functions the quadratic differential will have a pole of order $n+1$ at the point. He proved a coefficient result which represents a quite special case of the principle but did not obtain any general result of this type.

The General Coefficient Theorem was presented originally as an explicit embodiment of Teichmüller's principle, that is, the competing functions were subjected to the normalizations implied by the above statement [1, 2] Later improvements have made possible certain relaxations of these normalizations and we now state the result in the following form.

GENERAL COEFFICIENT THEOREM. *Let $\mathcal{R}$ be a finite Riemann surface, $Q(z)dz^2$ a positive quadratic differential on $\mathcal{R}$, $\{\Delta\}$ an admissible family of domains Δ_j, $j=1,\ldots,K$, on $\mathcal{R}$ relative to $Q(z)dz^2$ and $\{f\}$ an admissible family of functions $f_j, j=1,\ldots,K$, associated with $\{\Delta\}$. Let $Q(z)dz^2$ have double poles $P_1,\ldots,P_r$ and poles $P_{r+1},\ldots,P_n$ of order greater than two. We allow either of these sets to be void but not both. Let $P_j, j\leqslant r$, lie in the domain Δ_l and in terms of a local parameter z representing P_j as the point at infinity let f_l have the expansion*

$$f_l(z)=a^{(j)}z+a_0^{(j)}+\text{negative powers of } z \tag{1}$$

and Q the expansion

$$Q(z)=\alpha^{(j)}z^{-2}+\text{higher powers of } z^{-1}. \tag{2}$$

Let P_j, $j>r$, a pole of order m_j greater than two lie in the domain Δ_l and in terms of a local parameter z representing P_j as the point at infinity let f_l have the expansion

$$f_l(z)=z+\sum_{i=k_j}^{\infty}a_i^{(j)}/z^i, \tag{3}$$

where k_j is the smallest integer greater than or equal to $\frac{1}{2}m_j-2$ and Q the expansion

$$Q(z)=\alpha^{(j)}[z^{m_j-4}+\sum_{i=1}^{\infty}\beta_i^{(j)}z^{m_j-i-4}]. \tag{4}$$

Let

$$T_j=\alpha^{(j)}[a_{m_j-3}^{(j)}+\sum_{i=1}^{k_j}\beta_i^{(j)}a_{m_j-i-3}^{(j)}] \tag{5}$$

if m_j is odd and

$$T_j = \alpha^{(j)}[a^{(j)}_{m_j-3} + \sum_{i=1}^{k_j+1} \beta^{(j)}_i a^{(j)}_{m_j-i-3} + \tfrac{1}{2}(\tfrac{1}{2}m_j - 2)(a^{(j)}_{k_j})^2] \tag{6}$$

if m_j is even.

Then

$$\mathcal{R}\{\sum_{j=1}^{r} \alpha^{(j)} \log a^{(j)} + \sum_{j=r+1}^{n} \alpha^{(j)} T_j\} \leqslant 0, \tag{7}$$

where $\log a^{(j)} = \log |a^{(j)}| - id(F, P_j)$, $j \leqslant r$.

If equality occurs in (7) *each* f_j, $j = 1, \ldots, K$, *must be an isometry in the Q-metric*

$$|d\zeta| = |Q(z)|^{\frac{1}{2}} |dz|,$$

each trajectory of $Q(z)dz^2$ *in* $\bigcup_{j=1}^{K} \Delta_j$ *must go into another such and the set* $\bigcup_{j=1}^{K} f_j(\Delta_j)$ *must be dense in* $\mathcal{R}$. *If equality occurs in* (7) f_l *reduces to the identity in a domain* Δ_l *for which any of the following conditions holds.*

(i) *There is in* Δ_l *a pole* P_j, $j > r$, *of order* m_j *such that* $a^{(j)}_i = 0$ *for* $i <$ $\min(k_j + 1, m_j - 3)$.

(ii) *There is in* Δ_l *a pole* P_j, $j \leqslant r$, *with the corresponding coefficient* $a^{(j)}$ *equal to one.*

(iii) *There is in* Δ_l *a simple pole of* $Q(z)dz^2$ *or a point on a trajectory ending in a simple pole.*

Equality can occur in (7) *when there exists a double pole* P_j, $j \leqslant r$, *such that for the corresponding coefficient* $|a^{(j)}| \neq 1$ *only when* $\mathcal{R}$ *is conformally equivalent to the sphere and* $Q(z)dz^2$ *is a quadratic differential whose only critical points are two poles each of order two. If further* $\{\Delta\}$ *consists of a single domain the corresponding admissible function is conformally equivalent to a linear transformation with the points corresponding to these poles as fixed points.*

The present statement differs from the most general one previously given [3, 4, 5] in that in the latter the normalization $\beta^{(j)}_i = 0$, $i = 1, \ldots, k_j$, was imposed in the expansion (4) for $j = r+1, \ldots, n$. In [3] it was pointed out that this normalization was inessential in the sense that it could always be attained by a suitable change in the local parameter employed at the pole P_j. Thus it constituted no restriction in any application where poles of explicit numerical order were involved. However for theoretical purposes it is highly desirable to eliminate it as in the present formulation. This general form is quite readily obtained from the more special one by proving that the appropriate expression (5) or (6) is invariant under change of local parameter at P_j (always with the assumption that P_j be represented by the point at infinity) [6].

The terms finite Riemann surface and positive quadratic differential are employed with their standard meanings. We denote by C the set of zeros and simple poles of $Q(z)dz^2$, by H the set of poles of $Q(z)dz^2$ of order at least two. The trajectories of $Q(z)dz^2$ are the maximal regular curves on which $Q(z)dz^2 > 0$. An admissible family of domains relative to $Q(z)dz^2$ is obtained by slitting $\mathcal{R}$ along a finite number of trajectories of $Q(z)dz^2$ which either are closed or join two points of C and along a finite number of arcs on closures of trajectories in $\mathcal{R} - H$. An admissible family of functions then consists of a family of conformal mappings of the domains Δ_j onto non-overlapping domains in $\mathcal{R}$, leaving fixed all poles of $Q(z)dz^2$ interior to these domains, normalized by the expansions (3) at poles of order greater than two and admitting an admissible homotopy F into the identity such that

for a pole of order greater than two on the boundary of a strip domain the deformation degree of F at that pole is zero. As usual $d(F,P_j)$ denotes the deformation degree of F at the pole P_j.

The first, and perhaps most fundamental, step in the proof of the General Coefficient Theorem is to study the global structure of the trajectories of a positive quadratic differential. It is shown that there are five types of basic domains in this structure: end and strip domains (mapped respectively by $\int(Q(z))^{\frac{1}{2}}\,dz$ on half-planes and horizontal strips), circle and ring domains (mapped respectively by $\exp\,(k\int(Q(z))^{\frac{1}{2}}\,dz)$ for suitable constants k on circles and circular rings) and density domains which are swept out apart from possible points of C by trajectories each of which is everywhere dense in the domain.

Next it is shown that it is sufficient to prove the result under the assumption that $Q(z)dz^2$ has no simple poles on $\mathcal{R}$.

The proof culminates in drawing suitable curves around each point of H and regarding the set obtained by deleting these curves and their interiors from $\bigcup_{j=1}^{K}\Delta_j$. Then the areas of this set and the corresponding portion of $\bigcup_{j=1}^{K} f_j(\Delta_j)$ in the Q-metric are compared in two ways. First we obtain an inequality in the one direction by studying the behaviour of the functions f_j on the above curves. An inequality in the opposite direction is obtained by applying the method of the extremal metric in each of the various domains associated with the trajectory structure of $Q(z)dz^2$. Combining these inequalities we are led, after formal reduction, to the inequality (7).

The equality statements are proved by applying the usual equality treatment associated with the method of the extremal metric.

Now there are a variety of results, of which a prototype is the classical Area Theorem, which make assertions on the coefficients of univalent functions without the imposition of any normalization on the coefficients in question. Of course on the whole these results are far less powerful and less widely applicable than the General Coefficient Theorem, even allowing for our recent observation that the range of application of the area principle can be greatly increased by using it in connection with a variety of metrics [7].

If we somewhat strengthen the requirements in the enunciation of the General Coefficient Theorem the same method as is used for its proof, with appropriate technical modifications, provides the following result in which all normalizations on the coefficients of the mapping functions are eliminated. This result does not, of course, supersede the General Coefficient Theorem, since in turn it does not imply some of the most interesting consequences of the latter.

SPECIAL COEFFICIENT THEOREM. *Let $\mathcal{R}$ be a finite Riemann surface. Let $Q(z)dz^2$ be a positive quadratic differential on $\mathcal{R}$ such that each branch of $\int(Q(z))^{\frac{1}{2}}\,dz$ is single-valued in a sufficiently small neighborhood of each pole of $Q(z)dz^2$ (with that pole deleted) of order greater than two. Let $\{\Delta\}$ be an admissible family of domains Δ_j, $j=1,\ldots,K$, on $\mathcal{R}$ relative to $Q(z)dz^2$ and $\{f\}$ a special admissible family of functions f_j, $j=1,\ldots,K$, associated with $\{\Delta\}$. Let $Q(z)dz^2$ have double poles $P_1,\ldots,P_r$ and poles $P_{r+1},\ldots,P_n$ of order greater than two. Let P_j, $j\leqslant r$, lie in the domain Δ_l and in terms of a local parameter z representing P_j as the point at infinity let f_l have the expansion*

$$f_l(z)=a^{(j)}z+a_0^{(j)}+\text{negative powers of } z$$

and Q the expansion

$$Q(z)=\alpha^{(j)}z^{-2}+\text{higher powers of } z^{-1}.$$

For $j>r$ let ζ denote a specifically chosen branch of $\int(Q(z))^{\frac{1}{2}}\,dz$ in a neighborhood of P_j. Let $\gamma(P_j,L)$ denote the antecedent on $\mathcal{R}$ of the trace on the Riemann image of a neighborhood of P_j under ζ of a square in the ζ-plane of side $2L$, centre at the origin and with sides parallel to the real and imaginary axes, where L is to be sufficiently large that $\gamma(P_j,L)$ bounds a simply-connected neighborhood $U(P_j,L)$ of P_j. The curve $\gamma(P_j,L)$ is to be sensed so that $U(P_j,L)$ lies to its right. Let $P_j, j>r$, lie in the domain Δ_l and let ω denote the single-valued function obtained in a neighborhood of P_j by substituting $f_l(z)$ for z in the function ζ. Let Φ denote the intersection of $\bigcup_{j=1}^{K}\Delta_j$ with the union of density domains in the trajectory structure of $Q(z)dz^2$ and let $\Theta_j(L)=\Delta_j-\Phi-\bigcup_{k=r+1}^{n}\bar{U}(P_k,L)$, $j=1,\ldots,K$. Then

$$\sum_{i=1}^{K}\iint_{\Theta_i(L)}\left|(Q(f_i(z)))^{\frac{1}{2}}f_i'(z)-(Q(z))^{\frac{1}{2}}\right|^2 dA_z$$
$$\leqslant -\mathcal{R}\Big(2\pi\sum_{j=1}^{r}\alpha^{(j)}\log a^{(j)}\Big)+\sum_{j=r+1}^{n}\left(\frac{1}{2i}\int_{\gamma(P_j,L)}(\bar{\omega}-\bar{\zeta})\,d(\omega-\zeta)-\mathcal{J}\int_{\gamma(P_j,L)}\omega\,d\zeta\right)+o(1).$$

Here $\log a^{(j)}=\log|a^{(j)}|-id(F,P_j)$, $j\leqslant r$, where d denotes the deformation degree. The roots $(Q(f_i(z)))^{\frac{1}{2}}$, $(Q(z))^{\frac{1}{2}}$ correspond under passage from $f_i(z)$ to z along the path curve of the deformation F. Finally dA_z denotes the element of area for the local parameter z.

A special admissible family of functions and its special admissible homotopy F into the identity differ from the customary admissible family of functions and admissible homotopy into the identity in that for the former zeros of odd order play much the same role as simple poles do for the latter.

In various special circumstances this result reduces to a generalized Area Theorem, for example, if the family $\{\Delta\}$ consists of a single simply-connected domain which contains no simple poles or zeros of odd order of $Q(z)dz^2$ and $Q(z)dz^2$ has no poles of order two, or if $\mathcal{R}$ is the z-sphere, $Q(z)dz^2$ has a single pole and no zeros of odd order and $\{\Delta\}$ consists of a single domain.

REFERENCES

[1]. Jenkins, James A., A general coefficient theorem. *Trans. Amer. Math. Soc.* 77 (1954), 262–280.

[2]. —— *Univalent Functions and Conformal Mapping.* Springer-Verlag, Berlin–Göttingen–Heidelberg, 1958.

[3]. —— An extension of the General Coefficient Theorem. *Trans. Amer. Math., Soc.* 95 (1960), 387–407.

[4]. —— The General Coefficient Theorem and its applications. *Contributions to Function Theory,* 211–218. Tata Institute of Fundamental Research, Bombay, 1960.

[5]. —— The General Coefficient Theorem and certain applications, *Bull. Amer. Math. Soc.,* 68 (1962), 1–9.

[6]. —— An addendum to the General Coefficient Theorem. *Trans. Amer. Math. Soc.* (To appear.)

[7]. —— Some area theorems and a special coefficient theorem (To appear.)

РАЗЛИЧНЫЕ ПОДХОДЫ К ОЦЕНКЕ ТРУДНОСТИ ПРИБЛИЖЕННОГО ЗАДАНИЯ И ВЫЧИСЛЕНИЯ ФУНКЦИЙ

А. Н. КОЛМОГОРОВ

Чтобы избежать технических осложнений мы будем рассматривать только действительные функции

$$y = f(x)$$

заданные на отрезке

$$\Delta = \{x: -1 \leqslant x \leqslant +1\}.$$

Все наиболее существенные черты проблематики, которая нас будет занимать, можно понять на примере функциональных классов W_p — функций удовлетворяющих неравенству

$$\text{vrai max}\, |f^{(p)}(x)| \leqslant 1.$$

A_r — функций, которые могут быть аналитически продолжены во внутренность эллипса с фокусами в точках -1 и $+1$ и суммой полуосей $r > 1$.

Мы будем изучать способы задания и вычисления функций указанных классов с точностью до некоторого фиксированного $\varepsilon > 0$.

1. Приближение алгебраическими многочленами

Будем обозначать через $n_a(f, \varepsilon)$ наименьшее n, при котором существует многочлен

$$P(x) = c_1 + c_2 x + \ldots + c_n x^{n-1}$$

удовлетворяющий на Δ неравенству

$$|f - P| \leqslant \varepsilon$$

а для функционального класса F положим

$$n_a(F, \varepsilon) = \sup_{f \in F} n_a(f, \varepsilon).$$

Вот наиболее окончательные результаты об ассимптотическом поведении функций $n_a(\varepsilon)$ для классов W_p и A_r, которые дает классическая теория наилучших приближений:

$$n_a(W_p, \varepsilon) \sim C_p \left(\frac{1}{\varepsilon}\right)^{1/p}, \tag{1}$$

где C_p некоторые константы, аналитическое выражение которых мы приводить не будем (Никольский, Бернштейн, 1946–47);

$$n_a(A_r, \varepsilon) \sim \frac{1}{\log r} \log \frac{1}{\varepsilon} \tag{2}$$

(Бернштейн, 1913) (1).

2. Приближение линейными формами

Естественно поставить вопрос о том, насколько удачен выбор алгебраических многочленов в качестве аппарата для приближенного представления функций классов W_p и A_r (кажется, проблематика этого рода была впервые выдвинута в [3]). Обозначим через $n(F,\varepsilon)$ наименьшее n, при котором можно найти такие функции

$$\varphi_1, \varphi_2, \ldots, \varphi_n,$$

что для любой функции $f \in F$ найдется линейная форма

$$L(x) = c_1 \varphi_1(x) + c_2 \varphi_2(x) + \ldots + c_n \varphi_n(x)$$

удовлетворяющая на Δ неравенству

$$|f - L| \leqslant \varepsilon.$$

Легко понять, что всегда

$$n(F, \varepsilon) \leqslant n_a(F, \varepsilon).$$

В случае класса A_r

$$n(F, \varepsilon) \sim n_a(F, \varepsilon) \sim \frac{1}{\log r} \log \frac{1}{\varepsilon} \tag{3}$$

(Витушкин, Ерохин, 1957–58).

В случае класса W_p алгебраические многочлены оказываются не рациональным способом приближения, так как

$$n(F, \varepsilon) \sim C'_p \left(\frac{1}{\varepsilon}\right)^{1/p}, \tag{4}$$

где

$$\frac{C'_p}{C_p} = \frac{2}{\pi} \tag{5}$$

(Тихомиров, 1960).

---

(1) Современное изложение проблематики наших §§ 1–2 см. [1]. Вместо (2) в [1] сообщаются найденные позднее обобщения. В более старомодной формулировке (2) можно найти уже в § 46 классического мемуара Бернштейна [2]. В (1), (2) и далее

$n(\varepsilon) \sim m(\varepsilon)$, если $n(\varepsilon)/m(\varepsilon) \to 1$ при $\varepsilon \to 0$; $n(\varepsilon) \succeq m(\varepsilon)$, если $\liminf\limits_{\varepsilon \to 0} n(\varepsilon)/m(\varepsilon) > 0$;

$n(\varepsilon) \asymp m(\varepsilon)$, если одновременно $n(\varepsilon) \succeq m(\varepsilon)$ и $m(\varepsilon) \succeq n(\varepsilon)$.

Иначе говоря, выбирая рационально функции φ_k можно при той же точности ε ограничиться формами содержащими в

$$\frac{\pi}{2}=1{,}57\ldots$$

раза меньше членов, чем при приближении многочленами.

3. Необходимое количество информации

Чтобы фиксировать с точностью до ε функцию f, принадлежащую к классу F, в котором все функции допускают представление линейными формами

$$L(x)=c_1\varphi_1(x)+c_2\varphi_2(x)+\ldots+c_n\varphi_n(x)$$

с ошибкой не превышающей

$$|f-L|\leqslant\frac{\varepsilon}{2}$$

достаточно указать коэффициенты l_k с ошибкой не большей

$$\varepsilon'=\frac{\varepsilon}{2nM},$$

где M верхняя грань абсолютных величин функций φ_k.

Приближенные значения коэффициентов можно записывать по двоичной системе счисления. Таким образом, задание функции сводится к заданию определенного числа знаков, каждый из которых может принимать лишь значения 0 и 1. Если ограничиться функциями, подчиненными дополнительному условию

$$|f(x)|\leqslant 1,$$

то при приближении функций классов W_p и A_r алгебраическими многочленами при $\varepsilon\to 0$ число знаков, с которым необходимо задавать коэффициенты оказывается порядка $\log 1/\varepsilon$,

Это позволяет получить следующие оценки «количества информации», необходимого для фиксации с точностью до ε произвольной функции f класса A_r, или класса W_p:

$$I(A_r,\varepsilon)\preceq\left(\log\frac{1}{\varepsilon}\right)^2, \tag{6}$$

$$I(W_p,\varepsilon)\preceq\left(\frac{1}{\varepsilon}\right)^{1/p}\log\frac{1}{\varepsilon}. \tag{7}$$

Первая из этих оценок не может быть усилена в смысле порядка величины $I(A_r,\varepsilon)$, но может быть уточнена:

$$I(A_r, \varepsilon) \sim \frac{1}{2 \log r} \left(\log \frac{1}{\varepsilon}\right)^2 \tag{8}$$

(Витушкин, 1957).

Оценка же (7) дает немного завышенный порядок величины $I(W_p, \varepsilon)$. Истинный порядок таков:

$$I(W_p, \varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{1/p} \tag{9}$$

(Колмогоров 1955).

Более подробно о величине $I(f, \varepsilon)$ см. во введении к статье [4]. Там показывается, что $I(f, \varepsilon)$ есть целое число определяемое неравенствами

$$\log_2 N(F, \varepsilon) \leqslant I(F, \varepsilon) < \log_2 N(F, \varepsilon) + 1, \tag{10}$$

где $N(F, \varepsilon)$ минимальное число элементов покрытия множества F множествами диаметра 2ε в метрике

$$\varrho(f_1, f_2) = \sup_{x \in \Delta} |f_1(x) - f_2(x)|.$$

4. Сложность вычисления

Для простоты ограничимся в этом разделе функциями, которые кроме условия

$$|f(x)| \leqslant 1$$

введенного в предшествующем разделе, подчинены условию Липшица

$$|f(x'') - f(x')| \leqslant 1.$$

Пусть
$$s \geqslant \log \frac{1}{\varepsilon} + 3.$$

Легко видеть, что любая таблица, которая каждому двоичному числу

$$x = -1 + x_1, x_2 \ldots x_s \quad (x_i = 0,1)$$

ставит в соответствие двоичное число

$$y = -1 + y_1, y_2 \ldots y_s = f_\varepsilon(x), \quad (y_i = 0,1) \tag{11}$$

удовлетворяющее условию

$$|f(x) - f_\varepsilon(x)| \leqslant \frac{\varepsilon}{2}$$

может рассматриваться как таблица, дающая значения $y = f(x)$ с точностью до ε для любого $x \in \Delta$.

Зависимость (11) может интерпретироватьая как дискретная векторная функция

$$Y=\Phi(X)$$

отображающая s-мерные векторы

$$X=(x_1, x_2, \ldots, x_s)$$

в s-мерные векторы

$$Y=(y_1, y_2, \ldots. y_s),$$

где переменные x_j и y_i принимают лишь значения 0 и 1.

Такого рода векторные «функции алгебры логики» естественно представлять в виде суперпозиций элементарных функций

$$\zeta=f(\xi, \eta),$$

где воспомогательные переменные ξ, η, ζ принимают лишь значения 0, 1. Известно, что таких элементарных функций имеется всего 16. Будем называть «сложностью» векторной функции алгебры логики $Y=\Phi(X)$ минимальное число членов в представлении ее в виде суперпозиции элементарных функций (см. [5]).

Сложностью ε-задания функции f мы будем называть минимальную сложность функций Φ, соответствующий таблицам, задающим $y=f(x)$ с точностью до ε.

Соединяя результаты заметок [5] и [6] с некоторыми приемами, разработанными Шенноном, и оценками, изложенными в разделе 3 Офман установил, что максимальные ε-сложности функций из W_p и A_r допускают оценки

$$K(W_p, \varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{1/p} \Big/ \log \frac{1}{\varepsilon} \tag{12}$$

$$\left(\log \frac{1}{\varepsilon}\right)^2 \Big/ \log \log \frac{1}{\varepsilon} \preceq K(A_r, \varepsilon) \preceq \left(\log \frac{1}{\varepsilon}\right)^{\alpha}, \ \alpha=\log_2 3+1=2{,}58\ldots \tag{13}$$

Незавершенность оценок (13) связана с тем обстоятельством, что до настоящего времени, повидимому, остается неизвестной ассимптотика числа $m(s)$ элементарных операций над парами двоичных знаков, необходимого для выполнения умножения двух s-значных чисел. Обычный школьный рецепт умножения дает лишь

$$m(s) \preceq s^2$$

а оценки Офмана и Карацубы таковы:

$$s \preceq m(s) \preceq s^{\log_2 3}. \tag{14}$$

5. Оценка сложности индивидуальных функций

В классической теории приближений алгебраическими многочленами известны «обратные теоремы», показывающие, что ни одна индивидуальная функция, не обладающая достаточной «гладкостью», не может хорошо приближаться многочленами не слишком высоких степеней на всем отрезке Δ.

Все сказанное в разделах 2 и 3 по существу относится к классам функций, а не к индивидуальным функциям. Но при подходе к делу, изложенному в разделе 4, задача оценки ε-сложностей $K(f,\varepsilon)$ индивидуальных функций содержательна и очень интересна. К сожалению, получение оценок снизу здесь, видимо, очень трудно. Нарушение гладкости не делает еще неизбежным образом функцию сложной. Например, для функции Ван-дер Вардена, не имеющей ни в одной точке производной, легко получается оценка

$$K(f,\varepsilon) \preceq \left(\log \frac{1}{\varepsilon}\right)^2. \tag{15}$$

ЛИТЕРАТУРА

[1]. Тихомиров, В. М., *УМН*, 15, 3 (1960), 81–120.
[2]. Bernstein, S., *Bull. Acad. Roy. Belg.* ser. 2, 4 (1913).
[3]. Kolmogoroff, A., *Ann. Math.*, 37 (1936), 107–111.
[4]. Колмогоров, А. Н., и Тихомиров, В. М., *УМН*, 14, 2 (1959), 3–86.
[5]. Офман, Ю., *ДАН* 145, 1 (1962).
[6]. Карацуба, А., и Офман, Ю., *ДАН*, 145, 2 (1962).

ON DEFORMATIONS OF COMPACT COMPLEX STRUCTURES

By MASATAKE KURANISHI

By a holomorphic family of compact complex structures we mean a triple $(\mathcal{V},\mathcal{M},\tilde{\omega})$, where $\mathcal{V}$ and $\mathcal{M}$ are complex analytic spaces and $\tilde{\omega}$ is a surjective holomorphic mapping of $\mathcal{V}$ to $\mathcal{M}$, such that, for each t in $\mathcal{M}$, $\tilde{\omega}^{-1}(t)$ is a non-singular compact complex submanifold, say $V(t)$, of $\mathcal{V}$. We assume further that they satisfy certain supplementary conditions, which we will not discuss here. $\mathcal{M}$ will be referred to as the parameter space of the family. Thus the family gives rise to a collection $\{V(t);\ t\in\mathcal{M}\}$ of compact complex manifolds which can be considered as depending holomorphically on t in the parameter space. In the following, we may write $\{V(t);\ t\in\mathcal{M}\}$ to denote a holomorphic family under the understanding that we have a definite way of putting complex structure $\mathcal{V}$ on the set $\cup\{V(t);\ t\in\mathcal{M}\}$. Similarly we can consider C^∞ (or real analytic) family of compact complex structures. But we will restrict ourselves to holomorphic families in this lecture. If we fix a compact complex manifold V and take a family $\{V(t);\ t\in\mathcal{M}\}$ with a reference point t_0 in $\mathcal{M}$ such that $V(t_0)=V$, we may consider it as deforming the given structure V depending on parameter t. For this reason such a family will be called a (holomorphic) family of deformations of V with reference point t_0. Then for each tangent vector L to $\mathcal{M}$ at t_0, we can define what amount to the infinitesimal deformation of the above deformation to the direction L, and it turns out that this infinitesimal deformation, say $\tau(L)$, is in $H^1(V,\Theta)$, where Θ is the sheaf of germs of local holomorphic vector fields on V. The mapping $T_{t_0}(\mathcal{M})\ni L\to\tau(L)\in H^1(V,\Theta)$ is linear and will be called the infinitesimal deformation of the given family. When $\mathcal{M}$ is non-singular at t_0, the family is called effective at t_0 when τ is injective.

For families of compact complex manifolds of dimension 1 we have a very well developed theory as was exposed by the excellent lecture of Professor Ahlfors at this conference. In this case we have deep results concerning the local aspect of the families (i.e. how we can deform given V on a sufficiently small neighborhood of the reference point in the parameter space) as well as results concerning the global aspect of the families. For the families of compact complex manifolds of arbitrary dimension the works so far done are mainly concerned with the local aspect of the families, though there are some works concerning the global aspect. As was pointed out above, the infinitesimal deformations of a given compact complex manifold V are in $H^1(V,\Theta)$. So a problem naturally arises to see how $H^1(V,\Theta)$ controls locally the deformations of V. The first important result in this direction was obtained in 1957 by Frölicher and Nijenhuis [1]. They discovered the above-mentioned connection between families of deformations of V and $H^1(V,\Theta)$ and proved the following: If $H^1(V,\Theta)=0$, any family of deformations of V is locally trivial at the reference point, i.e. $(\mathcal{V},\mathcal{M},\tilde{\omega})$ is homeomorphic to $(V\times\mathcal{M},\mathcal{M},\tilde{\omega}_0)$ where $\tilde{\omega}_0$ is the canonical projection, provided we restrict $\mathcal{M}$ to a small neighborhood of the reference point. Kodaira and Spencer also took up the subject and made an extensive study in [2, 3]. One

of the important notions in their research is the notion of induced families. Namely, if $F=\{V(t); t\in\mathfrak{M}\}$ is a holomorphic family, $\mathfrak{M}'$ is an analytic space, and if f is a holomorphic mapping of $\mathfrak{M}'$ to $\mathfrak{M}$, then $\{\mathfrak{V}(f(s)); s\in\mathfrak{M}'\}$ can be naturally considered as a holomorphic family of complex structures which is called the induced family from F by f. A family F of deformations of V is called universal for V when any family of deformations of V with reference point s_0 is homeomorphic to the induced family from F by a mapping which sends the reference point of F to s_0, provided we restrict ourselves to a small neighborhood of s_0. Among other results, Kodaira and Spencer showed many examples of V for which an effective and universal family exists and calculated the dimensions of the parameter spaces of such families.

When one tries to construct a family of deformations of V, an approach is to start from a linear mapping of $\{t; |t|<1\}$ into $H^1(V,\Theta)$. This is the linear part of the family we are trying to construct. When we try to construct the second order part we find that we have an obstruction which is in $H^2(V,\Theta)$. When this obstruction vanishes we can construct the second order part. Then again we find an obstruction for the third order part in $H^2(V,\Theta)$, and this process goes on. Thus when $H^2(V,\Theta)=0$, there is no serious difficulty in constructing formally a family of deformations of V by this method. But there was a serious difficulty in proving that the above construction can be made so that it actually converges to a family of deformations of V. This difficulty was overcome by Kodaira, Nirenberg, and Spencer in [4] by applying Hölder estimate in the theory of strongly elliptic partial differential equations. In this way they proved that there exists a universal and effective family of deformations of V provided $H^2(V,\Theta)=0$. Moreover the dimension of the parameter space of this family is equal to that of $H^1(V,\Theta)$. By developing their technique further, the author was able to prove in [5] the existence of a universal and, so to speak, effective family of deformations of any compact complex manifold. Moreover, the parameter space of such a family for V, say $\mathfrak{M}(V)$, can be considered as the analytic set of zeros of a holomorphic mapping h of a neighborhood of the origin in $H^1(V,\Theta)$ into $H^2(V,\Theta)$, and might have singular points. The latter fact was already observed in [2] by an example. The existence of the above mapping h shows that $\dim H^1(V,\Theta)-\dim H^2(V,\Theta)\leqslant\dim \mathfrak{M}(T)<\dim H^1(V,\Theta)$.

All the known examples of V for which $\dim \mathfrak{M}(V)\lneqq\dim H^1(V,\Theta)$ are the case when $\dim V\geqslant 3$. If $\dim V$ is 1, we have the equality. So the problem of determining if $\dim \mathfrak{M}(V)=\dim H^1(V,\Theta)$ for any compact complex manifold $\mathfrak{V}$ of complex dimension 2 is one of the outstanding open problems. Recently D. Mumford found an interesting example [6]: namely, he constructed a family $F=\{V(t); t\in\mathfrak{M}\}$, where $\mathfrak{M}$ is non-singular, such that F is effective and universal for every $V(t)$ and $\dim \mathfrak{M}(V(t))=\dim \mathfrak{M}\lneqq\dim H^1$ $H^1(V(t),\Theta(t))$ for every t in $\mathfrak{M}$. This is the first example for which the generic member of an effetive and universal family has the inequality. He takes a suitable curve $C(t)$ imbedded in 3-dimensional projective space P and sets $V(t)$ to be the monoidal transform of P with center $C(t)$. It is interesting to note that all the known examples for which we have the inequality have something special about them: namely, they are either fibered manifolds or monoidal transforms of manifolds. So the problem of determing if $\dim \mathfrak{M}(V)=\dim H^1(V,\Theta)$, provided we exclude V which has certain special structures, is still open.

Let us mention some known theorems on the family of compact complex structures. When V is a Kählerian manifold and $\{V(t);\ t\in\mathfrak{M}\}$ is a family of deformations of V with reference point t_0, then $V(t)$ is also Kählerian for t sufficiently near t_0 [3]. However, Hironaka constructed in [7] an example of a family $\{V(t);\ t\in\mathfrak{M}\}$ such that $V(t)$ is Kählerian except one point in $\mathfrak{M}$, say t_1, and $V(t_1)$ is actually non-Kählerian. So the property of having a Kählerian metric is locally stable but not globally stable under deformations. Another theorem which has wide implication is the theorem of upper semi-continuity of cohomology groups: namely, if $\{V(t);\ t\in\mathfrak{M}\}$ is a holomorphic family and if $G(t)$ is a holomorphic family of coherent sheaves on $V(t)$ then $\dim H^a(V(t),G(t))$ is a upper semi-continuous function of t [3, 8] and $\{t;\ \dim H^a(V(t),G(t))\underset{\neq}{\leq} d\}$, where $d=\mathrm{Min}(\dim H^a(V(t),G(t)))$, is an analytic subset of $\mathfrak{M}$[8]. It should be mentioned that some of the known results about families of compact complex manifold were generalized to the case of families of non-compact complex manifolds by Grauert [8]. We note also that theory parallel to the above was also developed to some extent for the families of pseudo-group compact structures provided the pseudo-group satisfies certain conditions [9, 10].

REFERENCES

[1]. Frölicher, A. & Nijenhuis, A., A theorem on stability of complex structures. *Proc. Nat. Acad. Sci., U.S.A.*, 43 (1957), 239–241.

[2]. Kodaira, K. & Spencer, D. C., On deformations of complex analytic structures. I, II. *Ann. Math.*, 67 (1958), 328–466.

[3]. —— On deformations of complex analytic structures. III. *Ann. Math.*, 71 (1960), 43–76.

[4]. Kodaira, K., Nirenberg, L. & Spencer, D. C., On the existence of deformations of complex analytic structures. *Ann. Math.*, 68 (1958), 450–459.

[5]. Kuranishi, M., On the locally complete families of complex analytic structures. *Ann. Math.*, 75 (1962), 536–577.

[6]. Mumford, D., A farther pathology in algebraic geometry. (To appear in *Amer. J. Math.*)

[7]. Hironaka, H., An example of a non-Kählerian complex-analytic deformation of Kählerian complex structures. *Ann. Math.*, 75 (1962), 190–208.

[8]. Grauert, H., Der analytischen Garbentheorie und die Modulräume komplexer Strukturen. *Pub. Math.*, (*I.H.E.S.*), *Paris*, 5 (1960).

[9]. Kodaira, K., On deformations of some complex pseudogroup structures. *Ann. Math.*, 71 (1960), 224–302.

[10]. Spencer, D. C., Deformation of structures on manifolds defined by transitive continuous pseudogroups. (To appear in *Ann. Math.*)

PROLONGEMENT DE LA TRANSFORMATION DE LAPLACE

Par JEAN LERAY

La primitive $I(x,y)$ d'ordre $n-1$, s'annulant $n-1$ fois en y, d'une fonction $F(x)$ de la variable x est donnée par la formule de Cauchy :

$$I(x,y)=\int_y^x \frac{(x-t)^{n-2}}{(n-2)!}F(t)\,dt.$$

Introduisons dans cette formule les fonctions linéaires de x

$$\xi\cdot x=\xi_0+\xi_1 x, \quad \text{où } \xi=(\xi_0,\xi_1);$$

notons f la fonction homogène de degré $-n$

$$f(\xi)=\frac{1}{\xi_1^n}F\left(-\frac{\xi_0}{\xi_1}\right),$$

$\omega^*(\xi)$ la forme différentielle

$$\omega^*(\xi)=\xi_0\,d\xi_1-\xi_1\,d\xi_0$$

et h un arc joignant les ensembles de points ξ d'équations respectives

$$y^*:\ \xi\cdot y=0, \quad x^*:\ \xi\cdot x=0;$$

il vient
$$I(x,y)=\int_h \frac{(\xi\cdot x)^{n-2}}{(n-2)!}f(\xi)\,\omega^*(\xi).$$

Nous allons étudier l'intégrale analogue en un nombre quelconque de variables; nous obtiendrons ainsi un prolongement $\mathcal{L}$ de la transformation de Laplace, qui permet de construire la solution élémentaire d'opérateurs différentiels.

1. *La fonction* $\mathcal{L}[f]$. — Soit $\tilde{\Psi}$ une multiplicité analytique complexe, de dimension complexe

$$\dim_c \tilde{\Psi}=l;$$

ses points sont notés $\tilde{\varphi}$. Soient $\tilde{\Sigma}(x,y)$ un sous-variété analytique de $\tilde{\Psi}$, de codimension 1, d'équation réelle, dépendant *linéairement* de x et *holomorphiquement* de y; x est un point d'un espace affine réel X, ayant même dimension que $\tilde{\Psi}$:

$$\dim X=l;$$

y est un point d'un domaine Y de X.

Pour écrire l'équation de $\tilde{\Sigma}$, notons $(x_1,\ldots,x_l)$ les coordonnées de $x\in X$ et

$$\xi \cdot x = \xi_0 + \xi_1 x_1 + \ldots + \xi_l x_l$$

une fonction linéaire, numérique complexe, sur X; ξ est donc un vecteur d'un espace vectoriel complexe Ξ;

$$\dim_c \Xi = l+1.$$

Notons ξ^* l'image de $\xi \neq 0$ dans l'espace projectif

$$\Xi^* = (\Xi - 0)/\Lambda$$

quotient de $\Xi - 0$ par le groupe Λ de ses homothéties; autrement dit : ξ^* est l'hyperplan de X d'équation $\xi \cdot x = 0$.

L'équation locale de $\tilde{\Sigma}(x, y)$ est donc

$$\tilde{\Sigma}(x, y): \ \xi(\tilde{\varphi}, y) \cdot x = 0,$$

$\xi(\tilde{\varphi}, y)$ étant une application holomorphe d'un domaine de $\tilde{\Psi} \times Y$ dans Ξ; deux choix différents de $\xi(\tilde{\varphi}, y)$ sont proportionnels; nous supposerons que

$$\xi(\tilde{\varphi}, y) \neq 0;$$

ces divers $\xi(\tilde{\varphi}, y)$ définissent donc une application holomorphe de $\tilde{\Psi} \times Y$ tout entier :

$$\xi^*(\tilde{\varphi}, y): \tilde{\Psi} \times Y \to \Xi^*;$$

nous la nommerons *projection*; elle projette $\tilde{\Sigma}(x,y)$ (et $\tilde{\Sigma}(y,y)$) dans les hyperplans x^* (et y^*) de Ξ^* d'équations :

$$x^*: \ \xi \cdot x = 0, \quad y^*: \xi \cdot y = 0.$$

Nous notons $K(y)$ l'ensemble des points x de Y tels que $\tilde{\Sigma}(x,y)$ ait une singularité.

Nous notons $\omega^*(\xi)$ la forme différentielle extérieure

$$\omega^*(\xi) = \textstyle\sum_{j=0}^{l} (-1)^j \xi_j \, d\xi_0 \wedge \ldots \wedge d\xi_{j-1} \wedge d\xi_{j+1} \wedge \ldots \wedge d\xi_l;$$

son produit par une fonction de ξ, homogène de degré $-l-1$, est évidemment une forme différentielle de ξ^*.

Soit $f(\xi, y)$ une fonction numérique complexe, homogène en ξ, de degré $-n$ (n : entier >0, <0 ou $=0$), analytique en (ξ, y), en général multiforme et définie seulement sur une partie de $\Xi \times Y$. Nous dirons que *la projection* $\xi^*(\varphi, y)$ *uniformise la fonction* $f(\xi, y)$ (*ou la forme* $f(\xi, y)\omega^*(\xi)$) quand

$$f(\xi(\tilde{\varphi}, y), y) \quad (\text{ou } f(\xi(\tilde{\varphi}, y)\omega^*(\xi(\tilde{\varphi}, y)))$$

est une fonction (ou forme) de $(\tilde{\varphi}, y)$ holomorphe sur le domaine de $\tilde{\Psi} \times Y$ où $\xi(\tilde{\varphi}, y)$ est holomorphe et non nul; si $n=0$ (ou $=l+1$), cette fonction (ou forme) dépend de $\xi^*(\tilde{\varphi}, y)$ et $f(\xi, y)$ sans dépendre du choix de $\xi(\tilde{\varphi}, y)$, qu'on fait au voisinage de chaque point.

Supposons que $\tilde{\Sigma}(y,y)$ contienne une sous-multiplicité $\tilde{\Sigma}_1$ de $\tilde{\Psi}$, indépen-

dante de y; supposons $\tilde{\Sigma}(x,y)$ et $\tilde{\Sigma}_1$ en position générale pour $x \neq y$. Supposons donnée

si $n \leqslant l$, une classe d'homologie $h(\tilde{\Sigma}, \tilde{\Sigma}_1)$ de $\tilde{\Sigma}$ relativement à $\tilde{\Sigma}_1$, ayant la dimension $l-1$;

si $n > l$ une classe d'homologie $h(\tilde{\Psi}, \tilde{\Sigma} \cup \tilde{\Sigma}_1)$ de dimension l et notons alors $h(\tilde{\Sigma}, \tilde{\Sigma}_1)$ son bord dans $\tilde{\Sigma}$:

$$\partial h(\tilde{\Psi}, \tilde{\Sigma} \cup \tilde{\Sigma}_1) = h(\tilde{\Sigma}, \tilde{\Sigma}_1). \tag{1}$$

Supposons que ces classes dépendent continûment(1) de (x, y).

Définissons la fonction $\mathcal{L}[f]$ de (x, y), pour $y \in Y$ et $x \in Y - K(y)$, par l'intégrale:

$$\mathcal{L}[f] = \frac{1}{(2\pi i)^{l-1}} \frac{1}{2} \int_{h(\tilde{\Psi}, \tilde{\Sigma} \cup \tilde{\Sigma}_1)} \frac{(\xi \cdot x)^{n-l-1}}{(n-l-1)!} f(\xi, y)\, \omega^*(\xi) \quad \text{si } n > l, \tag{2.1}$$

$$= \frac{(-1)^{n-l-1}}{(2\pi i)^{l-1}} \frac{1}{2} \int_{h(\tilde{\Sigma}, \tilde{\Sigma}_1)} \frac{d^{l-n}[f(\xi \cdot y)\, \omega^*(\xi)]}{(d\xi \cdot x)^{1+l-n}} \quad \text{si } n \leqslant l, \tag{2.2}$$

où $\xi^* = \xi^*(\tilde{\varphi}, y)$; on suppose $f\omega^*$ uniformisé par cette projection, la seconde intégrale porte sur une classe-résidu (voir : [7], III).

La formule (2.1) généralise (1) puisque les projections de $\tilde{\Sigma}$ (et $\tilde{\Sigma}_1$) sont dans x^*[et dans y^*]. Cette fonction $\mathcal{L}[f]$ de (x, y) sera notée parfois $\mathcal{L}[f]\,(x, y)$.

Voici ses *propriétés* :

Pour x et $y \in Y$, $x \notin K(y)$, $\mathcal{L}[f]$ est une fonction *holomorphe* de (x, y).

Si les $c_i(y)$ sont des fonctions holomorphes de y et si une même projection uniformise les $f_i\omega^*$, alors

$$\sum_i c_i(y) \mathcal{L}[f_i] = \mathcal{L}[\sum_i c_i f_i]. \tag{3}$$

Quand $f\omega^*$ est uniformisé, alors

$$\frac{\partial}{\partial x_j} \mathcal{L}[f] = \mathcal{L}[\xi_j f]\,; \left[n - l - 1 - \sum_j x_j \frac{\partial}{\partial x_j}\right] \mathcal{L}[f] = \mathcal{L}[\xi_0 f]. \tag{4}$$

Quand f est uniformisé, alors

$$x_j \mathcal{L}[f] = -\mathcal{L}\left[\frac{\partial f}{\partial \xi_j}\right], \quad \mathcal{L}[f] = -\mathcal{L}\left[\frac{\partial f}{\partial \xi_0}\right] \tag{5}$$

$$\frac{\partial}{\partial y_j} \mathcal{L}[f] = \mathcal{L}\left[\frac{\partial f}{\partial y_j}\right]. \tag{6}$$

On a
$$\mathcal{L}[f(\Theta\xi, y)]\,(x, y) = t^{l-n} \mathcal{L}[f(\xi, y)]\,(Tx, y), \tag{7}$$

(1) Quand (x,y) varie suffisamment peu, elles contiennent un cycle variant continûment.

T désignant l'homothétie de centre y, de rapport t :

$$T : x \to y + t(x-y)$$

et Θ la transformation, opérant sur Ξ et Ξ^* :

$$\Theta : \xi = (\xi_0, \xi_1, \dots, \xi_l) \to (t\xi \cdot y - \xi_1 y_1 - \dots - \xi_l y_l, \xi_1, \dots, \xi_l),$$

qui vérifie $$\Theta\xi \cdot Tx = t\xi \cdot x.$$

Voici deux conséquences évidentes de ces propriétés :

Si $(\xi \cdot y)^{p+1} f(\xi)$ est uniformisé et si $f(\xi)$ est indépendant de ξ_0, alors

$$\mathcal{L}\left[\frac{(-\xi \cdot y)^p}{p!} f(\xi)\right] = \mathcal{L}\left[\frac{(-\xi \cdot y)^{p+1}}{(p+1)!} f(\xi)\right]$$

ne dépend que de $x-y$ et est homogène en $x-y$ de degré $n-l$. Si $f(\xi)$ est un polynome, alors la *fonction* $\mathcal{L}[f]$ est nulle; cependant la *distribution* $\mathcal{L}[f]$, que nous allons définir, ne le sera pas : elle aura pour support $x=y$.

2. *La distribution* $\mathcal{L}[f]$ *pour* $x \neq y$. — Faisons maintenant les hypothèses suivantes :

$$y \in K(y) \ ;$$

pour $x \in K(y) - y$, $\tilde{\Sigma}(x, y)$ a un seul point singulier, $\tilde{\sigma}(x, y)$; $\tilde{\sigma}$ appartient à une partie compacte de $\tilde{\Psi}$; $\tilde{\sigma} \notin \tilde{\Sigma}_1$; $\tilde{\sigma}$ est un point double quadratique de $\tilde{\Sigma}$, c'est-à-dire, en notant Hess le Hessien :

$$\xi(\tilde{\varphi}, y) \cdot x = 0, \ \frac{\partial \xi(\tilde{\varphi}, y)}{\partial \tilde{\varphi}} \cdot x = 0, \ \mathrm{Hess}_{\tilde{\varphi}} [\xi(\tilde{\varphi}, y) \cdot x] \neq 0 \text{ pour } \tilde{\varphi} = \tilde{\sigma}(x, y).$$

Ces hypothèses ont les conséquences suivantes : $K(y) - y$ est une *sous-multiplicité* de Y, enveloppe des hyperplans de Y d'équation $\xi(\tilde{\varphi}, y) \cdot x = 0$ (y fixe, $\tilde{\varphi}$ variable); nous noterons l'équation locale ou cette sous-multiplicité :

$$K(y) - y : \ k(x, y) = 0.$$

La partie singuliére de $\mathcal{L}[f]$, sur $K(y)$, est

$$H_0 k^{n-1-l/2} + H_1 \log k \quad (l \text{ pair} \geqslant 2n),$$

$$H_0 k^{n-1-l/2} \log k \quad (l \text{ pair} < 2n),$$

$$H_0 k^{n-1-l/2} \quad (l \text{ impair}),$$

où $H_i = H_i(x, y)$ est holomorphe; la restriction de H_0 à K se calcule sans quadrature.

L'étude de cette partie singulière définit sans ambiguité sur $X - y$ une *distribution* de x, fonction de y, $\mathcal{L}[f]$; sa restriction à $X - K(y)$ est la fonction $\mathcal{L}[f]$; cette distribution $\mathcal{L}[f]$ possède les propriétés qu'énonce le n° 1.

3. *Le conoïde* $K(y)$. — Complétons l'hypothèse que $\tilde{\Sigma}(y, y)$ contient une sous-multiplicité $\tilde{\Sigma}_1$ de $\tilde{\Psi}$. Notons $\tilde{\Gamma}$ l'ensemble des points singuliers de

$\tilde{\Sigma}(y, y)$ appartenant à $\tilde{\Sigma}_1$; nous le supposerons non vide[1] et $\neq \tilde{\Sigma}_1$; faisons alors les hypothèses les plus simples possibles :

1° $\tilde{\Sigma}_1$ *est l'espace projectif complexe*, de $\dim_c : l-1$;

2° $\tilde{\Sigma}_1 \cap \tilde{\Sigma}(x, y)$ *est un hyperplan de* $\tilde{\Sigma}_1$, arbitraire quand x décrit un voisinage de y; par contre $\tilde{\Gamma}$ n'est pas un hyperplan;

3° $\tilde{\Psi}$ est un voisinage de $\tilde{\Sigma}_1$ dans *un espace fibré* $\tilde{\Phi}$ *ayant* $\tilde{\Sigma}_1$ *pour section* et la droite complexe pour fibre.

D'après H. Grauert, son article [4] prouve, vu 1° et 2°, que 3° équivaut à

3°bis $\tilde{\Sigma}_1$ est *exceptionnelle* dans $\tilde{\Psi}$; c'est-à-dire : remplacer $\tilde{\Sigma}_1$ par un point, en conservant la structure analytique de $\tilde{\Psi} - \tilde{\Sigma}_1$, transforme $\tilde{\Psi}$ en un espace analytique.

Voici les *conséquences* de ces hypothèses :

D'après la théorie des espaces fibrés analytiques, on a

$$\tilde{\Phi} = (\Phi - \Phi_0)/\Lambda,$$

où : Φ est l'espace vectoriel complexe de $\dim_c : l+1$;
$(\tau, \eta_1, \ldots, \eta_l)$ désignera les coordonnées de $\varphi \in \Phi$;
Φ_0 est le sous-espace : $\eta_1 = \ldots = \eta_l = 0$ de Φ;
Λ est le groupe de transformations à un paramètre λ :
$\lambda : \ \varphi = (\tau, \eta_1, \ldots, \eta_l) \to \lambda\varphi = (\lambda^{1-m}\tau, \lambda\eta_1, \ldots, \lambda\eta_l)$ (m : entier).

La projection $\qquad \xi^*(\tilde{\varphi}, y) : \tilde{\Psi} \times y \to \Xi^*$

est induite par une application commutant avec λ :

$$\xi(\varphi, y) : \ \Psi \times Y \to \Xi \quad \text{où} \quad \Psi/\Lambda = \tilde{\Psi} \, ; \, \xi(\lambda\varphi, y) = \lambda\xi(\varphi, y).$$

Les équations de $\tilde{\Sigma}_1$ et $\tilde{\Sigma}(x, y) \cap \tilde{\Sigma}_1$ sont

$$\tilde{\Sigma}_1 : \tau = 0 \, ; \quad \tilde{\Sigma}(x, y) \cap \tilde{\Sigma}_1 : \ \tau = \eta_1(x_1 - y_1) + \ldots + \eta_l(x_l - y_l) = 0.$$

$\xi^*(\tilde{\varphi}, y)$ projette homéomorphiquement $\tilde{\Sigma}_1$ sur y^* : on peut faire en sorte que si $\varphi = (0, \eta_1, \ldots, \eta_l)$, alors $\xi(\varphi, y) = \eta = (\eta_0, \eta_1, \ldots, \eta_l)$ vérifiant $\eta \cdot y = 0$.

Le déterminant fonctionnel

$$\left| \frac{D(\xi_0(\varphi, y), \ldots, \xi_l(\varphi, y))}{D(\tau, \eta_1, \ldots, \eta_l)} \right|_{\tau=0} = g(\eta, y)$$

est un polynome en $\eta_1, \ldots, \eta_l$, homogène de degré m; on nomme g : polynome de la projection $\xi^*(\tilde{\varphi}, y)$; l'équation de $\tilde{\Gamma}$ est donc :

$$\tilde{\Gamma} : \tau = g(\eta, y) = 0 \, ;$$

donc $m > 1$ (ce qui permet de voir que 1°, 2° et 3° impliquent 3 bis). La projection de $\tilde{\Gamma}$ sur y^* est la variété algébrique :

---

(1) Afin de pouvoir définir des classes d'homologie non nulles.

$$g^* : \ \eta \cdot y = g(\eta, y) = 0.$$

Nous supposerons $\quad \operatorname{Hess} g(\eta, y) \neq 0 \quad$ sur[1] $\operatorname{Re} g^*$.

Alors : $K(y)$ *est un conoïde de sommet* y; ses plans tangents η^* en son point conique y sont les $\eta^* \in \operatorname{Re} g^*$.

Le support des singularités de $f(\xi, y)$ *détermine* $K(y)$, à l'exception de celles de ses nappes correspondant aux nappes de $\operatorname{Re} g^*$ sur lesquelles $f(\xi, y)$ est holomorphe.

4. *Choix*[2] *de* $h(\tilde{\Psi}, \tilde{\Sigma} \cup \tilde{\Sigma}_1)$.—Supposons la projection $\xi^*(\tilde{\varphi}, y)$ *hyperbolique*, c'est-à-dire le polynome $g(\eta, y)$ hyperbolique : quand $\eta_2, \ldots, \eta_l$ sont réels, non tous nuls, alors les η_1 annulant g sont réels et distincts.

Définissons d'abord une classe d'homologie $h(y^* - g^*, x^*)$:

si l *est impair* et $x_1 \geqslant y_1$, c'est celle de $2 \operatorname{Re} y^*$, muni d'une orientation qui change sur $x^* \cap y^*$, détourné[3] de g^*;

si l *est pair* et $x_1 \geqslant y_1$, c'est celle du cobord[4] de $\operatorname{Re} g^*$, muni d'une orientation qui change sur $g^* \cap x^*$.

On constate que $\quad h(y^* - g^*, x^*) = 0 \quad$ pour $x_1 = y_1$ $\qquad$ (8)

et on définit $\quad h(y^* - g^*, x^*) = 0 \quad$ pour $x_1 < y_1$. $\qquad$ (9)

Note 4. Si l est *impair* on peut ne pas supposer la projection hyperbolique; on n'aura plus ni (8), ni (9); on choisira toujours $h(y^* - g^*, x^*)$ classe de $2 \operatorname{Re} y^*$ détourné de g^*; ce qui suit vaudra à condition de remplacer $\frac{1}{2}$ par $\frac{1}{4}$ au second membre de (2).

L'homéomorphisme de y^* et $\tilde{\Sigma}_1$ transforme $h(y^* - g^*, x^*)$ en une classe $h(\tilde{\Sigma}_1 - \tilde{\Gamma}, \tilde{\Sigma})$; rappelons que $\tilde{\Sigma}_1 \cap \tilde{\Sigma}$ ne dépend que de la direction de $x - y$. On peut alors définir $h(\tilde{\Sigma}, \tilde{\Sigma}_1)$ par la condition de dépendre continûment de (x, y) pour $x \neq y$ et d'être voisin[5] de $h(\tilde{\Sigma}_1 - \tilde{\Gamma}, \tilde{\Sigma})$ quand x tend vers y, le long d'un segment non tangent à $K(y)$.

On définit enfin $h(\tilde{\Psi}, \tilde{\Sigma} \cup \tilde{\Sigma}_1)$ par la condition (1).

Ce choix a les conséquences suivantes :

$\mathcal{L}[f] = 0$ *pour* $x_1 < y_1$, si g est hyperbolique;

$\mathcal{L}[f] = 0$ pour $x \neq y$ signifie que f est un polynome;

$\mathcal{L}[f]$ *ne dépend pas du choix de la projection* $\xi^*(\tilde{\varphi}, y)$ *qui uniformise* f; *si* $1 + (l/2) < n$, $\mathcal{L}[f]$ *est une fonction* $O(|x - y|^{n-l})$ pour $x - y$ petit.

5. *La distribution* $\mathcal{L}[f]$. Supposons $f(\xi, y)$ *rationnellement uniformisable;* c'est-à-dire : quel que soit le polynome homogène $b(\xi)$ de $(\xi_1, \ldots, \xi_2)$, il existe une projection $\xi^*(\tilde{\varphi}, y)$ uniformisant $f(\xi, y)/b(\xi)$, ayant pour polynome $g(\eta, y) b(\eta)$, g étant indépendant de b. Supposons f hyperbolique, c'est-à-dire g hyperbolique (ou bien l impair : voir Note 4).

---

(1) Re...: partie réelle de ....

(2) L. Gårding a contribué à ce choix.

(3) *Détourner* d'une sous-variété analytique complexe des segments réels orientés la coupant, c'est remplacer chaque segment par la demi-somme de deux demi-circonférences, imaginaires conjuguées l'une de l'autre, ayant ce segment pour diamètre, ayant son orientation, ne rencontrant pas la sous-variété.

(4) Prendre *le cobord* de points d'une sous-variété analytique, c'est remplacer chaque point par une circonférence enlacée avec cette sous-variété.

(5) Contenir un cycle voisin d'un cycle choisi dans ....

Il existe alors une distribution $\mathcal{L}[f]$ de $x \in Y$, fonction de $y \in Y$ et possédant toutes les propriétés énoncées ci-dessus : vu (4) on la définit par la formule

$$\mathcal{L}[f] = b\left(\frac{\partial}{\partial x}\right) \mathcal{L}[f(\xi, y)/b(\xi)],$$

où b est hyperbolique de degré $>1+(l/2)-n$, $\mathcal{L}[f/b]$ est une fonction et $b\ \mathcal{L}[f/b]$ est une distribution indépendante du choix de b.

$\mathcal{L}[f]=0$ implique $f=0$; $\mathcal{L}[1]$ est défini et vaut

$$\mathcal{L}[1] = \delta(x-y) \quad (\delta : \text{mesure de Dirac}); \tag{10}$$

plus généralement, *quand $f(\xi)$ est une fonction rationnelle* homogène de $(\xi_1, \ldots, \xi_l)$ à dénominateur hyperbolique, $\mathcal{L}[f]$ est l'expression de sa transformée de Laplace classique qu'ont donnée Herglotz [5] et Petrowsky [8].

Note. Rappelons que Gel'fand, Shapiro, Shilov [3], Gårding [2] ont récemment étudié les transformées de Fourier des fonctions homogènes.

6. *La solution élémentaire $E(x,y)$* d'un opérateur différentiel $a(y, \partial/\partial y)$, linéaire, d'ordre m et hyperbolique si l est pair est

$$E(x,y) = \mathcal{L}[U(\xi,y)] \tag{11}$$

à condition que

$$a\left(y, \frac{\partial}{\partial y}\right) U(\xi, y) = 1, \tag{12}$$

$U(\xi,y)$ et ses dérivées en y d'ordres $<m$ soient rationnellement uniformisables. (13)

Or (13) implique que :

$$U(\xi,y) \text{ s'annule } m \text{ fois pour } \xi \cdot y = 0; \tag{14}$$

et la solution du problème de Cauchy (12), (14) possède la propriété (13); voir [7] II.

La formule (11), ainsi obtenue, permet d'analyser les singularités de $E(x,y)$ et aussi de calculer explicitement $E(x,y)$ dans le cas suivant : les coefficients de a d'ordres m, $m-1$ et $<m-1$ sont repectivement linéaires, constants et nuls [7]. F. John [6] a, dans le cas elliptique, déjà employé (11) (l impair) et pu traiter le cas : l pair.

7. *Un prolongement de la convolution par une transformée de Laplace,* analogue au prolongement $\mathcal{L}$ de la transformation de Laplace, permet de résoudre de même le problème de Cauchy à données analytiques; voir [7] V; il se trouve, en partie, chez Fantappiè [1].

BIBLIOGRAPHIE

[1]. Fantappiè, L., L'indicatrice proiettiva dei funzionali lineari e i prodotti funzionali proiettivi. *Annali Mat. ser. IV*, 13 (1943).

—— *Second Colloque sur les Equations aux Dérivées Partielles* (Bruxelles, 1954), 95–128.

—— La théorie des équations aux dérivées partielles. *Colloque C.N.R.S.* (Nancy, 1956), 47–62.

[2]. GÅRDING, L., Transformation de Fourier des distributions homogènes. *Bull. Soc. Math. France*, 89 (1961).

[3]. GEL'FAND, I. M. & SHILOV, G. E., *La Théorie des Distributions Dunod*. 1961.

[4]. GRAUERT, H., Über Modifikationen und exceptionelle analytische Mengen. *Math. Ann.*, 146, (1962), 331–368.

[5]. HERGLOTZ, G., Über die Integration linearer, partieller Differentialgleichungen mit konstanten Koeffizienten. *Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math. phys.*, 78 (1926), 92–126 et 287–318; 80 (1928), 68–114.

[6]. JOHN, F., The fundamental solution of linear elliptic differential equations with analytic coefficients. *Comm. Pure Appl. Math.*, 3 (1950), 213–304.

[7]. LERAY, J., Problème de Cauchy. I. *Bull. Soc. math. France*, 85 (1957), 389–429; II. *ibid.*, 86 (1958), 75–96; III. *ibid.*, 87 (1959), 81–180; IV. *ibid.*, 90 (1962), 39–156; V. *C.R. Acad. Sc. Paris*, 242 (1956), 953–959; VI. *Cahiers de physique*, 133 (1961), 373–381.

[8]. PETROWSKY, I., On the diffusion of waves and the lacunas for hyperbolic equations. *Mat. Sb.* (Recueil math.), 17 (1945), 289–370.

ENSEMBLES DE RÉSOLUTION SPECTRALE

Par PAUL MALLIAVIN

Le problème de la synthèse spectrale de Beurling-Gel'fand est énoncé habituellement comme celui de la classification des sous-espaces fermés de $L_\infty(Z)$ invariants par translation.

Pour la commodité des notations, nous préférerons considérer ici le problème équivalent obtenu par transformation de Fourier.

On note par A l'algèbre de Banach des fonctions réelles définies sur le cercle Γ, ayant une série de Fourier absolument convergente.

P dénotera l'espace de Banach réel dual de A; P peut être identifié avec l'espace des distributions, au sens de Laurent Schwartz, définies sur Γ et ayant leurs coefficients de Fourier bornés. Si $T \in P$ on notera par $s(T)$ le support de la distribution T. On notera enfin par (a, T), $a \in A$, $T \in P$ la forme bilinéaire réalisant la dualité.

On peut immédiatement définir sur P une structure de A-module. Si $a_0 \in A$ et $T \in P$ sont donnés alors le produit de $a_0 T$ est l'élément de P défini par

$$(a, a_0 T) = (a a_0, T),$$

où a est un élément générique de A.

P sera muni de la topologie vague c'est-à-dire la topologie la moins fine rendant continues toutes les fonctions (a, T).

Ceci étant le problème de synthèse spectrale peut s'éconcer comme celui de la classification des sous-modules fermés de P. Plus précisément si $\mathfrak{M}$ est un tel sous-module on lui associe son spectre

$$\sigma(\mathfrak{M}) = \bigcup_{T \in \mathfrak{M}} s(T)$$

qui est une partie fermée de Γ. Une partie fermée E de Γ est un *ensemble de synthèse spectrale* s'il existe un *seul* sous-module fermé $\mathfrak{M}_E$ tel que

$$\sigma(\mathfrak{M}_E) = E.$$

On peut trouver dans un travail récent de Carl Herz [2] une étude exhaustive de toutes les classes des ensembles qui sont actuellement connus comme étant de synthèse spectrale.

La notion d'ensemble de synthèse spectrale n'est pas très maniable. On ne sait pas démontrer, par exemple, que l'union de deux ensembles de synthèse spectrale est encore de synthèse spectrale. Calderón [1] a introduit une sous-classe des ensembles de synthèse spectrale qui se conserve par l'opération de réunion. Nous procéderons de façon analogue.

On dit qu'une partie fermée F de Γ est un ensemble de *résolution spectrale* si toute partie fermée E de F est de synthèse spectrale. Si F est de résolution spectrale la structure des sous-modules fermés de $\mathfrak{M}_E$ est complètement connue : elle est isomorphe à l'algèbre de Boole des parties fermées de F.

On voit aisément que si F est fermé, $F=\bigcup_{j=1}^{+\infty} F_j$, les ensembles F_j étant tous de résolution spectrale, alors F est lui-même de résolution spectrale.

Utilisant un raisonnement de dualité élémentaire bien connu on a que E est un ensemble de synthèse spectrale si et seulement si

$$s(T)\subset E\subset a^{-1}(0) \quad (T\in P, a\in A)$$

entraîne
$$(a,T)=0. \tag{0.1}$$

Kahane et Salem ont construit [4] des ensembles parfaits H tels que

$$T\in P,\ s(T)\subset H \quad \text{entraîne que} \quad T \text{ est une mesure.}$$

Utilisant (0.1) on voit immédiatement qu'un tel ensemble H est un ensemble de résolution spectrale.

Nous nous proposons ici d'établir un lien entre les ensembles de résolution spectrale et une notion de la théorie classique des séries trigonométriques : celle d'ensemble d'unicité.

Rappelons qu'une partie G de Γ est *un ensemble d'unicité* si le fait pour une série trigonométrique de converger vers zéro sur le complémentaire de G entraîne que tous ses coefficients sont nuls. Si G est fermé on peut transformer cette définition de la manière suivante :

P_0 notera la partie de P constituée par les distributions dont le coefficient de Fourier de rang n tend vers zéro lorsque $n\to\pm\infty$.

Alors G est un ensemble d'unicité si et seulement si

$$T\in P_0,\ s(T)\subset G \quad \text{entraîne} \quad T=0.$$

THÉORÈME 0.3. *Tout ensemble de résolution spectrale est un ensemble d'unicité.*

Si E est une partie fermée de Γ, l'algèbre de restrictions des fonctions de A à E, soit A_E est définie comme le quotient de A par l'idéal J_E des fonctions nulles sur E.

L'espace de Banach dual de A_E, soit P^E s'identifie avec l'espace des distributions $T\in P$ telles que

$$s(T)\subset E,$$

$$(a,T)=0 \quad \text{pour tout } a\in J_E.$$

En particulier toutes les mesures de support contenues dans E sont des éléments de P^E. On dira que E est un *ensemble de multiplicité au sens strict* s'il existe $T\in P^E$ dont les coefficients de Fourier tendent vers zéro.

Sous des conditions générales sur E, Kahane et Katznelson [3, 6] (cf. la conférence de M. Kahane dans ces *Proceedings*) ont montré que les fonctions opérant sur A_E sont seulement les fonctions analytiques. Le résultat que nous allons donner comme conséquence immédiate du Théorème 0.3 est moins précis, mais ne rentre pas, actuellement semble-t-il, dans le cadre de résultats cités ci-dessus.

THÉORÈME 0.4. *Soit E un ensemble de multiplicité au sens strict, soit Φ une fonction opérant sur A_E, alors Φ appartient à la classe de fonctions indéfiniment dérivables*

$$C((n\log n)^n).$$

1. Avant de démontrer ces théorèmes nous allons introduire quelques notations. Si $a \in A$ on note par

$$a'(n) = \int a(x)\, e^{-inx} \frac{dx}{2\pi}.$$

Si $T \in P$ on note par $\qquad T'(n) = (e^{inx}, T).$

Avec ces conventions $\qquad (a, T) = \sum_n a'(n)\, T'(n).$

On note par $\qquad \|a\|_1' = \sum |a'(n)|$

$$\|T\|_\infty' = \sup |T'(n)|.$$

Enfin remarquons que $\qquad \|bT\|_\infty' \leqslant \|b\|_1' \|T\|_\infty'.$

On notera enfin par $A_{\log}$ l'ensemble des fonctions $a \in A$ telles que

$$\sum |a'(n)| \log \left| \frac{1}{a'(n)} \right| < \infty. \tag{1.1}$$

On a alors le lemme

LEMME 2. *Supposons que l'on ait trouvé $T \in P$, $T'(0) \neq 0$ et $a \in A_{\log}$ tels que posant*

$$p(u) = \|e^{iua} T\|_\infty',$$

on ait

$$\int_{-\infty}^{+\infty} p(u)\,(|u| + 1)\, du < \infty. \tag{2.1}$$

Alors $s(T)$ n'est pas un ensemble de résolution spectrale.

(On peut remarquer que $a \in A$ implique $e^{iua} \in A$ et par suite $p(u)$ est bien défini.)

Preuve. On sait que [8] l'hypothèse (1.1) entraîne qu'il existe une suite M_n telle que la classe de fonctions indéfiniment différentiables définies sur la droite $C(M_n)$ soit non quasi-analytique et que si $F \in C(M_n)$ et est à support compact on ait

$$\int_{-\infty}^{+\infty} \|e^{iua}\|_1' \, |\hat{F}(u)|\, du < \infty, \tag{2.2}$$

où $\hat{F}$ est la transformée de Fourier de F définie par

$$\hat{F}(u) = \int_{-\infty}^{+\infty} F(t)\, e^{-itu} \frac{dt}{2\pi}.$$

Il résulte de (2.2) que l'intégrale

$$\int_{-\infty}^{+\infty} e^{iua}\hat{F}(u)\,du$$

converge comme intégrale vectorielle dans A. On appellera $F(a)$ la valeur de cette intégrale. La convergence dans A implique que

$$F(a)\,T = \int_{-\infty}^{+\infty} (e^{iua}T)\,\hat{F}(u)\,du$$

converge comme intégrale vectorielle dans P. En particulier on aura

$$(F(a), T) = \int_{-\infty}^{+\infty} (e^{iua}T)'(0)\,\hat{F}(u)\,du \tag{2.3}$$

et

$$\|F(a)\,T\|'_\infty \leqslant \int_{-\infty}^{+\infty} p(u)\,|\hat{F}(u)|\,du. \tag{2.4}$$

Soit maintenant $f \in C(M_n)$, le support de f étant contenu dans $(-1, +1)$ et de plus

$$\int f(t)\,dt = 1.$$

Soit g la dérivée de f. Posons

$$k_r(t) = -r^2 g(rt).$$

Alors on a quelque soit r $\qquad |\hat{k}_r(u)| < |u|.$

Par suite utilisant (2.4) et (2.1) on obtient que

$$\|k_r(a)\;T\|'_\infty \quad \text{est uniformément bornée.}$$

En extrayant au besoin une sous-suite on peut se ramener au cas où

$$k_r(a)\;T \to U$$

la convergence étant une convergence vague dans P. On a

$$s(U) \subset \limsup\,(s(k_r(a)\;T)) = s(T) \cap a^{-1}(0). \tag{2.5}$$

D'autre part $\qquad (a, U) = \lim\limits_{r\to\infty} (a, k_r(a)\,T) = \lim\limits_{r\to\infty} (ak_r(a), T). \qquad (2.6)$

Nous allons évaluer cette limite en utilisant (2.3). En combinant (2.3) et (2.4) on obtient que

$$|(F(a), T)| < \int_{-\infty}^{+\infty} p(u)\,|\hat{F}(u)|\,du. \tag{2.7}$$

Nous noterons par $l(F)$ l'expression dont la valeur absolue figure au premier membre de cette inégalité. Alors $l(F)$ est une forme linéaire sur

$C(M_n)$, continue pour la norme L_2 en vertu de (2.7) et de (2.1). Il existe donc une fonction h de carré intégrable, dont le support est contenu dans l'ensemble des valeurs de a, telle que

$$(F(a), T) = \int F(t)\, h(t)\, dt. \tag{2.8}$$

Prenant en particulier $F(t) = e^{iut}$ on obtient

$$|\hat{h}(u)| < p(u)$$

ce qui avec (2.1) entraîne que $h(t)$ est une fois continûment différentiable.

Calculons (2.6) en utilisant (2.8), on obtient

$$(a, U) = \lim_r \int t k_r(t)\, h(t)\, dt = h(0). \tag{2.9}$$

On peut toujours se ramener au cas $h(0) \neq 0$. En effet

$$\int h(t)\, dt = (1, T) = T'(0) \neq 0.$$

Par suite il existe v_0 tel que $h(v_0) \neq 0$. Il suffit d'appliquer le raisonnement précédent à $a_1 = a - v_0$ pour avoir $h_1(0) = h(v_0) \neq 0$.

Comparant (2.5) et (2.9) on voit que

$$H = a^{-1}(0) \cap s(T) \tag{2.10}$$

n'est pas un ensemble de synthèse spectrale, et comme $H \subset s(T)$, $s(T)$ n'est pas un ensemble de résolution spectrale, d'où le lemme.

3. Nous allons démontrer le théorème (0.3) en montrant qu'un ensemble qui n'est pas d'unicité ne peut pas être un ensemble de résolution spectrale. Si E n'est pas un ensemble d'unicité, on peut trouver $T \in P_0$, $T \neq 0$ avec $s(T) \subset E$. En multipliant T par e^{imx} on peut se ramener au cas ou $T'(0) \neq 0$.

Nous associerons à T la fonction décroissante tendant vers zéro suivante

$$g(n) = \max_{|r| > n} |T'(r)|.$$

T étant donné nous prendrons

$$a(x) = \sum_k k^{-1-\varepsilon} \sin n_k x, \tag{3.1}$$

où $\varepsilon > 0$ est fixé et où $\{n_k\}$ sera une suite lacunaire que nous construirons de façon convenable. D'une manière plus précise la suite $\{n_k\}$ sera déterminée par le lemme

Lemme 4. *On peut trouver une suite croissante d'entiers $\{n_k\}$ telle que quelque soit les entiers positifs s et q on ait*

$$\sum_{k=1}^{s} k^2 n_k < \frac{1}{4} n_{s+1}, \tag{4.1$_s$}$$

$$\sum_{r=q+1}^{+\infty} (2q^2+1)^{r-q} g\left(\frac{1}{4} n_{r-1}\right) < e^{-q^2}. \tag{4.2$_q$}$$

Preuve. D'abord il est clair que si une suite $\{m_k\}$ satisfait l'une des inégalités (4.1$_s$) ou (4.2$_q$) toute suite extraite de $\{m_k\}$ satisfera encore la même inégalité. Cette remarque nous permet d'utiliser le procédé diagonal.

On déterminera d'abord une suite $\{n_k^0\}$ qui satisfait (4.1$_s$) quelque soit s, ce qui est évidemment possible car si les n_k^0 sont déterminés pour $k \leqslant s$, on pourra toujours choisir n_{s+1} assez grand pour que (4.1$_s$) soit vérifié.

$\{n_k^1\}$ sera une suite extraite de $\{n_k^0\}$ telle que (4.2$_1$) soit satisfait.

$\{n_k^q\}$ sera une suite extraite de $\{n_k^{q-1}\}$ de telle sorte que (4.2$_q$) soit satisfait. Enfin on prend

$$n_k = n_k^k$$

et le lemme sera satisfait.

5. Nous allons établir dans ce paragraphe un lemme combinatoire. Nous noterons par f une fonction définie sur les entiers positifs, prenant des valeurs entières, égale à zéro sauf en un nombre fini de points. Nous noterons par F la famille de toutes ces fonctions.

F_q notera la partie de F constituée par les f vérifiant

$$f(k) = 0 \quad \text{si} \quad k < q,$$

$$|f(k)| \leqslant q^2 \quad \text{si} \quad k \geqslant q.$$

A une fonction f et à la suite n_k on associe l'entier $n(f)$ défini par

$$n(f) = \sum_k n_k f(k).$$

Avec ces notations on a

LEMME 5.1. *Soit m un entier donné. A tout entier $r \geqslant q$ associons*

$$L_r = \left\{ f \in F_q;\ |m + n(f)| < \frac{1}{4} n_r \right\}.$$

Alors L_r contient au plus

$$1 + (2q^2+1)^{r-q} \quad \textit{éléments.}$$

Preuve. Soit f_0 tel que

$$|m + n(f_0)| = \min_{f \in F_q} |m + n(f)|$$

et notons par d ce minimum

$$|m + n(f)| \geqslant d \tag{5.2}$$

et d'autre part

$$|m + n(f)| = |m + n(f_0) + n(f - f_0)| \geqslant |n(f - f_0)| - d. \tag{5.3}$$

Soit s le plus petit entier tel que

$$d < \frac{n_s}{4}.$$

Si $j \geqslant q$, notons par

$$C_j = \{f \in F_q;\ f(j) \neq f_0(j) \quad \text{et} \quad f(k) = f_0(k) \quad \text{si} \quad k > j\}.$$

(4.1) entraîne que $\qquad |n(f - f_0)| > \frac{1}{2} n_j \quad \text{si} \quad f \in C_j.$

Par suite d'après (5.3)

$$|n(f) + m| \geqslant \tfrac{1}{4} n_j \quad \text{si} \quad j \geqslant s \ \text{et} \ f \in C_j.$$

Si $j < s$ nous utilisons (5.2) et nous obtenons

$$|m + n(f)| \geqslant \tfrac{1}{4} n_{s-1} \quad \text{pour tout } f \in C_j.$$

Par suite $\qquad C_j \cap L_r = \emptyset \quad \text{si} \quad j \geqslant s \quad \text{et} \quad j \geqslant r. \qquad (5.4)$

D'autre part on a $\qquad F_q = \{f_0\} \cup (\bigcup_{j=q}^{+\infty} C_j)$

et (5.4) entraine alors que

$$L_r \subset \{f_0\} \cup (\bigcup_{j=q}^{r} C_j).$$

Il est facile d'énumérer le nombre d'éléments figurant dans le second membre et ceci démontre le lemme.

6. Nous allons aborder l'évaluation de $p(u)$, nous allons montrer que pour u assez grand,

$$p(u) < \exp(-\eta u^{1/(1+\varepsilon)}) \quad \text{où } \eta \text{ est une constante positive.} \qquad (6.1)$$

Nous associerons d'abord à u l'entier q défini par

$$(q-1)^2 (4e)^{-1} \leqslant |u| < q^2 (4e)^{-1}. \qquad (6.2)$$

Nous allons utiliser quelques identités de la théorie des fonctions de Bessel. On a

$$\exp(iuk^{-1-\varepsilon} \sin n_k x) = D_k + E_k, \qquad (6.3)$$

où $\qquad D_k = \sum_{s=-q^2}^{q^2} J_s(uk^{-1-\varepsilon}) \exp(isn_k x),$

E_k étant la somme des termes de rang s vérifiant $|s| > q^2$.

On a en utilisant des estimées asymptotiques classiques de la théorie des fonctions de Bessel (cf. [7] par exemple)

$$\|D_k\|_1' + \|E_k\|_1' < 4(|u| + 1). \qquad (6.4)$$

D'autre part on peut prouver de même qu'il existe un nombre $\delta < 1$ tel que

$$|J_s(v)| < \delta \quad \text{quel que soit} \quad v \quad \text{réel}, \ \tfrac{1}{4} < |v| < \tfrac{1}{2}, \qquad (6.5)$$

Enfin en tenant compte de (6.2) on peut montrer que

$$|J_s(uk^{-1-\varepsilon})| < (4k^{1+\varepsilon})^{-|s|} \quad \text{si} \quad |s| > q^2. \qquad (6.6)$$

w notera un ensemble fini d'entiers $\geqslant q$. W_q notera l'ensemble de tous les w. On notera par w° le complémentaire de w dans l'ensemble des entiers $\geqslant q$. Avec ces notations on a

LEMME 7. *On a, pour* $|u| > 2\|T\|'_\infty + 1$, *quelque soit* $w \in W$,

$$\|T \prod_{k \in w^0} D_k\|'_\infty < \|T\|'_\infty + 1, \tag{7.1}$$

et

$$\|T \prod_{k=q}^{+\infty} D_k\|'_\infty < \|T\|'_\infty \exp(-\eta u^{1/(1+\varepsilon)}), \tag{7.2}$$

où η *est une constante numérique.*

Preuve. On a en développant le produit infini

$$\prod_{k=q}^{+\infty} D_k = \sum_{f \in F_q} C(f) \exp(in(f)\,x),$$

où

$$C(f) = \prod_{k=q}^{+\infty} J_{f(k)}(uk^{-1-\varepsilon}); \tag{7.3}$$

on a

$$(T \prod_{k=q}^{+\infty} D_k)'(m) = \sum_{f \in F_q} T'(m + n(f))\, C(f).$$

Utilisant les notations du lemme (5.1) le second membre est majoré en module par

$$\|T\|'_\infty\, C(f_0) + \sum_{r=q+1}^{+\infty} \sum_{f \in B_r} |T'(m + n(f))|\,|C(f)|,$$

où

$$B_r = L_r - L_{r-1}.$$

Il résulte immédiatement de (7.3) que $|C(f)| < 1$.

Par suite en utilisant encore (5.1)

$$\sum_{f \in B_r} < (2q^2 + 1)^{r-q} g(\tfrac{1}{4} n_{r-1}).$$

Utilisant (4.2_q) on a

$$\|T \prod_{k=q}^{+\infty} D_k\|'_\infty \leqslant \|T\|'_\infty \sup_{f \in F_q} |C(f)| + e^{-|u|}. \tag{7.4}$$

Tout le raisonnement précédent est valable si on remplace le produit infini, k variant de q à $+\infty$ par un produit analogue k variant dans w°, il suffit de remplacer dans (7.4) $C(f)$ par le produit infini (7.3) pris encore sur w°. Ainsi (7.4) établit (7.1).

Pour établir (7.2) il suffit de remarquer que

$$|C(f)| < \prod_{\frac{1}{4} < |u| k^{-1-\varepsilon} < \frac{1}{2}} |J_{f(k)}(uk^{-1-\varepsilon})|$$

et d'utiliser (6.5) pour obtenir

$$|C(f)| < \exp(-\eta u^{1/(1+\varepsilon)}).$$

8. *Démonstration de l'inégalité* (6.1)

$$e^{iua}T = b\prod_{k=q}^{+\infty}(D_k+E_k)\,T$$

où
$$b=\prod_{k=1}^{q-1}(D_k+E_k),$$

d'ou
$$p(u)<\|b\|_1'\,\|\prod_{k=q}^{+\infty}(D_k+E_k)\,T\|_\infty'.$$

Utilisant (6.4) on a

$$\|b\|_1'<\exp(\eta' u^{\frac{1}{2}}\log u)$$

et nous pouvons négliger ce facteur si $1/(1+\varepsilon)>\frac{1}{2}$.

Ainsi 6.1 sera une conséquence de la même évaluation pour

$$p_1(u)=\|T\prod_{k=q}^{+\infty}(D_k+E_k)\|_\infty';$$

or on a
$$\prod_{k=q}^{+\infty}(D_k+E_k)=\prod_{k=q}^{+\infty}D_k+\sum_{w\in W}\prod_{k\in w^0}D_k\prod_{k\in w}E_k,$$

d'où utilisant (7.1)

$$p_1(u)\leqslant\|T\prod_{k=q}^{+\infty}D_k\|_\infty'+(\|T\|_\infty'+1)\sum_{w\in W}\prod_{k\in w}\|E_k\|_1'.$$

Il est facile d'évaluer la dernière somme.

En effet elle est égale à

$$\prod_{k=q}^{+\infty}(1+\|E_k\|_1')-1$$

et elle a le même ordre de grandeur que

$$\sum_{k=q}^{+\infty}\|E_k\|_1',$$

somme qui est inférieur à $e^{-|u|}$ en vertu de (6.6) ce qui établit l'inégalité (6.1) et démontre le théorème (0.3).

9. *Démonstration du Théorème* 0.4. Notons par J_E l'idéal constitué par les fonctions de A qui sont nulles sur E. Sur A_E est munie de la norme d'algèbre de Banach quotient

$$\|a\|_{1,E}'=\inf\|b\|_1',\quad \text{où } (b-a)\in J_E.$$

Si $T\in P^E$ alors on a en tenant compte du fait que T est orthogonal à J_E

$$\sup_{\|a\|_{1,E}'=1}|(a,T)|=\sup_{\|b\|_1'=1}(b,T)=\|T\|_\infty'. \tag{9.1}$$

Soit maintenant Φ une fonction opérant sur A_E, soit h la fonction associée au couple a, T par la formule (2.8). Calculons

$$I(u)=\int \Phi(\xi)\, e^{iu\xi} h(\xi)\, d\xi.$$

D'après (2.8) on a $\qquad I(u)=(\Phi(a)\ e^{iua}, T)$

ou encore $\qquad I(u)=(\Phi(a),\ e^{iua}\ T),$

d'où en utilisant (9.1) et la définition de $p(u)$ donnée dans le lemme 2, on obtient

$$|I(u)|\leqslant \|\Phi(a)\|_{1,E}\, p(u)$$

Nous pouvons nous dispenser de l'hypothèse $a\in A_{\log}$ et prendre simplement $a\in A$. Ceci implique, comme dans [8], que $C(M_n)$ étant une classe non quasi-analytique donnée, on peut construire a tel que

$$\|\Phi(a)\|'_{1,E}<\infty$$

entraîne $\qquad h\,\Phi\in C(M_n). \qquad (9.2)$

Prenant en particulier $\Phi=1$ on obtient que $h\in C(M_n)$.

Nous noterons par Q l'ensemble des classes $\mathcal{C}$ non quasi-analytiques de fonctions différentiables qui sont fermées pour la division c'est-à-dire telles que

$$k\in\mathcal{C},\ k\neq 0 \quad \text{entraîne} \quad \frac{1}{k}\in\mathcal{C}.$$

Utilisant (9.2) on obtient que si Φ opère sur A_E alors

$$\Phi\in\bigcap_{\mathcal{C}\in Q}\mathcal{C}$$

Rudin a montré [10] que $\bigcap_{\mathcal{C}\in Q}\mathcal{C}$ est exactement la classe $C((n\log n)^n)$ ce qui démontre le théorème (0.4).

10. *Remarques.* Tous les ensembles de non synthèse spectrale du cercle actuellement connus ne sont pas des exemples d'unicité.(1) Ce résultat est évidemment faux sur le Tore à 2 dimensions (x_1,x_2). Il existe des ensembles de non synthèse portés par la droite $x_2=0$.

On montrerait que l'ensemble H introduit en (2.10) est de multiplicité en montrant que

$$U'(n)\to 0 \quad |n|\to\infty$$

ce qui est une conséquence de la formule

$$U=\int (e^{iua}T)\, u\, du$$

et due fait que $T\in P_0$.

(1) Ce problème a été résolu depuis indépendamment par Filippi d'une part et Kahane Katznelson d'autre part. Ils ont montré que l'ensemble parfait de Cantor E n'est pas de résolution spectrale; ceci est d'autant plus remarquable que l'on sait d'après un théorème de C. Herz que E est un ensemble de synthèse spectrale. Enfin Filippi a étendu le théorème 0.3 à un groupe abélien non discret arbitraire, Kahane Katznelson ont montré que sous les hypothèses du théorème 0.4 seules les fonctions analytiques opèrent sur A_E.

Signalons un résultat immédiat sur les séries de Taylor lacunaires : *soit E un ensemble qui n'est pas d'unicité, on peut construire une série de Taylor lacunaire absolument convergente $f(z)$ telle que $f(E)$ ait comme partie du plan omcplexe un intérieur non vide.*

La technique pour obtenir ce résultat consiste à montrer comme en (2.8) qu'il existe une fonction continue $h(t,v)$ des 2 variables réelles, t et v telle que

$$(F(f(e^{i\theta})), T) = \int F(t,v)h(t,v)\, dtdv.$$

Cette formule peut être démontrée utilisant une technique analogue à celle de Rudin [9]. Comme le support de h est contenu dans $f(E)$, on a le résultat.

J. P. Kahane, G. Weiss and M. Weiss [5] ont obtenu des résultats analogues mais plus précis pour des ensembles spéciaux du type de Cantor.

Le fait que l'ensemble E ne soit pas d'unicité ne semble pas être lié de façon intrinseèque au problème mais seulement à la technique utilisée.

BIBLIOGRAPHIE

[1]. Calderón, A., Ideals in Abelian group algebras. *Symposium on Harmonic Analysis and Related Integral Transforms.* Cornell University, 1956.

[2]. Herz, Carl S., The spectral theory of bounded functions. *Trans. Amer. Math. Soc.*, 94 (1960), 180–232.

[3]. Kahane, J. P. & Katznelson, Y., Sur la réciproque du théorème de Wiener-Levy. *C.R. Acad. Sci. Paris*, 248 (1959), 1279–1281.

[4]. Kahane, J. P. & Salem, R., Sur les ensembles linéaires ne portant pas de pseudo-mesures. *C.R. Acad. Sci. Paris*, 243 (1956), 1185–1187.

[5]. Kahane, J. P., Weiss, G., & Weiss, M., On gap Taylor series. (A paraître aux *Math. Scand.*)

[6]. Katznelson, Y., Calcul symbolique sur les algèbres de restriction. *C.R. Acad. Sci. Paris*, 254 (1962), 2514–2515, 2700–2702.

[7]. Malliavin, P., Impossibilité de la synthèse spectrale sur les groupes abéliens non compacts. *Publ. Math. (I.H.E.S.), Paris*, 2 (1959), 61–68.

[8]. —— Calcul symbolique et sous-algèbres de $L_1(G)$. *Bull. Soc. Math. France*, 8 (1959) 180–190.

[9]. Rudin, W., *Monographie sur l'Analyse Harmonique.* (A paraître.)

[10]. —— *On inverse closed algebras of differentiable functions.* (A paraître.)

RÉSULTATS RÉCENTS ET PROBLÈMES DE NATURE ALGÉBRIQUE EN THÉORIE DE L'APPROXIMATION

Par LEOPOLDO NACHBIN

Nous discuterons dans cet exposé certains résultats connus et quelques problèmes ouverts de la théorie de l'approximation des fonctions réelles, continues ou différentiables, dans le cas des algèbres et modules de telles fonctions. Nous ferons cette étude soit du point de vue de la théorie de Weierstrass, donc pour la convergence uniforme sur tout compact, soit du point de vue plus général de la théorie de Bernstein, c'est-à-dire de la convergence uniforme ponderée sur des parties non nécessairement compactes.

I. Théorie de Weierstrass

La théorie de Weierstrass des fonctions réelles continues trouve son expression dans le théorème classique de Weierstrass–Stone sur la structure des sous-algèbres fermées de l'algèbre topologique $\mathcal{C}(E)$ des fonctions réelles continues sur un espace uniformisable séparé E, munie de la topologie compacte. Ce théorème entraîne des conséquences sur la nature des sous-espaces vectoriels fermés de $\mathcal{C}(E)$, qui sont des modules sur des sous-algèbres fermées de $\mathcal{C}(E)$; ou, plus généralement, sur la nature des sous-espaces vectoriels fermés des sommes continues d'espaces vectoriels topologiques réels localement convexes separés au dessus de E, qui sont des modules sur des sous-algèbres fermées de $\mathcal{C}(E)$. On est aussi amené aux résultats suivants de synthèse spectrale. Soit $\mathcal{W}$ un espace localement convexe réel. Considérons une algèbre commutative $\mathcal{A}$ d'opérateurs continus sur $\mathcal{W}$, contenant l'opérateur identité. Supposons que $\mathcal{W}$ soit localement convexe sous $\mathcal{A}$ par rapport à la catégorie $\{\mathbf{R}\}$ des algèbres isomorphes à l'algèbre $\mathbf{R}$ des nombres réels (pour la terminologie voir Nachbin [11] et que les opérateurs de $\mathcal{A}$ soient semblablement bornés (ce qui signifie que la topologie de $\mathcal{W}$ est déterminée par l'ensemble des semi-normes continues sur $\mathcal{W}$ par rapport à chacune desquelles tous les opérateurs de $\mathcal{A}$ sont bornés). Alors, tout sous-espace vectoriel fermé $\mathcal{A}$-invariant de $\mathcal{W}$ maximal est nécessairement de codimension 1. Tout sous-espace vectoriel fermé propre $\mathcal{A}$-invariant de $\mathcal{W}$ est contenu dans un sous-espace vectoriel fermé $\mathcal{A}$-invariant de $\mathcal{W}$ maximal; et, plus précisement, il est l'intersection de tels sous-espaces maximaux le contenant. De plus, $\mathcal{V}$ étant un sous-espace vectoriel $\mathcal{A}$-invariant de $\mathcal{W}$ et $\mathcal{A}'$ étant l'algèbre d'opérateurs induite par $\mathcal{A}$ sur l'espace localement convexe quotient $\mathcal{W}'=\mathcal{W}/\mathcal{V}$, $\mathcal{W}'$ est alors localement convexe sous $\mathcal{A}'$ par rapport à la catégorie $\{\mathbf{R}\}$ et les opérateurs de $\mathcal{A}'$ sont semblablement bornés. Dans les propositions précédentes, les conclusions sont fausses si l'on omet l'hypothèse que les opérateurs de $\mathcal{A}$ sont semblablement bornés. La recherche d'une hypothèse plus générale que celle ci, pour que ces propo-

sitions restent vraies, nous amène à la théorie de Bernstein. On obtient aussi des résultats analogues dans le cas des algèbres topologiques réelles localement convexes par rapport à la catégorie $\{\mathbf{R}\}$. La théorie de Weierstrass des fonctions réelles continuement différentiables n'est pas encore achevée même sur $\mathbf{R}^n$, la différence entre ce cas et celui des variétés différentiables étant banale. D'autre part, on ne voit pas encore clairement la forme que prendra cette théorie sur des espaces différentiables, généralisant les variétés différentiables et $\mathbf{R}^n$ en ce même sens ou les espaces uniformisables séparés généralisent les variétés topologiques et $\mathbf{R}^n$. Finalement, la forme de la théorie en termes de synthèse spectrale des algèbres commutatives d'opérateurs continus sur des espaces localement convexes et de synthèse spectrale des algèbres topologiques localement convexes n'est pas non plus éclaircie. Rappelons des résultats connus et quelques problèmes. Plaçons nous sur $\mathbf{R}^n$, pour simplifier l'exposé. Soit $\mathcal{E}^m(\mathbf{R}^n)$ l'algèbre topologique des fonctions réelles m-fois continuement différentiables sur $\mathbf{R}^n$, munie de la topologie compacte d'ordre m, où $0 \leqslant m < \infty$ (les modifications à introduire dans le cas $m = \infty$ étant évidentes). La structure des idéaux fermés de $\mathcal{E}^m(\mathbf{R}^n)$ est connue (voir Whitney [18]; voir aussi Schwartz [17]). On peut en formuler les résultats sous plusieurs formes. Disons que $f, g \in \mathcal{E}^m(\mathbf{R}^n)$ sont m-tangentes au point $x \in \mathbf{R}^n$ si leurs derivées d'ordre $\leqslant m$ au point x sont égales. $\mathcal{J}$ étant un idéal et f un élément de $\mathcal{E}^m(\mathbf{R}^n)$, pour que f soit adhérent à $\mathcal{J}$ dans $\mathcal{E}^m(\mathbf{R}^n)$ il faut et il suffit que, pour tout $x \in \mathbf{R}^n$, il existe $g \in \mathcal{J}$, m-tangente à f au point x. Ceci se reformule comme suit. Pour tout $x \in \mathbf{R}^n$, soit $\mathcal{J}_x^{m+1}$ l'idéal fermé de $\mathcal{E}^m(\mathbf{R}^n)$ des fonctions m-plates au point x, c'est à dire m-tangentes à la fonction 0 au point x. L'adhérence d'un idéal $\mathcal{J}$ de $\mathcal{E}^m(\mathbf{R}^n)$ est l'intersection des idéaux fermés $\mathcal{J} + \mathcal{J}_x^{m+1}$ pour $x \in \mathbf{R}^n$. En introduisant sur $\mathcal{E}^m(\mathbf{R}^n)$ la topologie simple d'ordre m, on voit que tout idéal fermé pour la topologie compacte d'ordre m est aussi fermé pour la topologie simple d'ordre m. C'est clair que tout idéal fermé maximal est de la forme $\mathcal{J}_x = \{f \in \mathcal{E}^m(\mathbf{R}^n);\ f(x) = 0\}$ pour un $x \in \mathbf{R}^n$ unique, donc de codimension 1. Tout idéal fermé propre de $\mathcal{E}^m(\mathbf{R}^n)$ est contenu dans un idéal fermé maximal; et, plus précisement, il est l'intersection des idéaux fermés élémentaires le contenant (Un idéal fermé propre $\mathcal{J}$ est dit élémentaire si l'ensemble ordonné des idéaux fermés contenant $\mathcal{J}$ strictement a un plus petit élément: le successeur de $\mathcal{J}$). Un idéal fermé élémentaire est toujours primaire (Un idéal fermé propre $\mathcal{J}$ est dit primaire si l'ensemble ordonné des idéaux fermés propres contenant $\mathcal{J}$ a un plus grand élément : l'unique idéal fermé maximal contenant $\mathcal{J}$). Un idéal fermé élémentaire $\mathcal{J}$ satisfait à $\mathcal{J}_x^{m+1} \subset \mathcal{J} \subset \mathcal{J}_x$ pour un $x \in \mathbf{R}^n$ unique. Tout idéal fermé élémentaire étant de codimension finie bornée par une valeur $c(m,n)$, on en déduit que tout idéal fermé de $\mathcal{E}^m(\mathbf{R}^n)$ est l'intersection des idéaux fermés le contenant et de codimension au plus égale à $c(m,n)$. Un autre résultat intéressant et plus récent, obtenu à partir de l'analyse des idéaux fermés de $\mathcal{E}^m(\mathbf{R}^n)$, est celui d'après lequel, pour $m = \infty$, donc pour $\mathcal{E}(\mathbf{R}^n)$ muni de la topologie compacte d'ordre infini, tout idéal engendré par un nombre fini de fonctions analytiques-réelles est nécessairement fermé (voit Łojasiewicz [4]; voir aussi Malgrange [5]). La nature des sous-algèbres denses de $\mathcal{E}^m(\mathbf{R}^n)$ est connue (voir Nachbin [10]; voir aussi Mirkil [11]). Pour qu'une partie $\mathcal{X}$ de $\mathcal{E}^m(\mathbf{R}^n)$ engendre une sous-algèbre dense de $\mathcal{E}^m(\mathbf{R}^n)$, il faut et il suffit que : (1) si $x, y \in \mathbf{R}^n$, $x \neq y$, il existe $f \in \mathcal{X}$ telle que $f(x) \neq f(y)$; (2) si $x \in \mathbf{R}^n$, il existe $f \in \mathcal{X}$ telle que $f(x) \neq 0$; (3) si $x \in \mathbf{R}^n$ et $\nu \in \mathbf{R}^n, \nu \neq 0$, il existe $f \in \mathcal{X}$ telle que

$$\frac{\partial f}{\partial \nu}(x) \neq 0.$$

(Pour $m=0$, la troisième condition est à omettre. Pour $m \geqslant 1$, ces conditions de densité sont indépendantes de m.) On peut reformuler ce résultat de plusieurs façons. $\mathcal{A}$ étant une sous-algèbre de $\mathcal{E}^m(\mathbf{R}^n)$, pour que $\mathcal{A}$ soit dense dans $\mathcal{E}^m(\mathbf{R}^n)$ il faut et il suffit que $\mathcal{A}$ vérifie la condition biponctuelle suivante : si $f \in \mathcal{E}^m(\mathbf{R}^n)$, $x, y \in \mathbf{R}^n$, il existe $g \in \mathcal{A}$, m-tangente à f aux points x et y. Une sous-algèbre de $\mathcal{E}^m(\mathbf{R}^n)$ est dense pour la topologie compacte d'ordre m si et seulement si elle l'est pour la topologie simple d'ordre m; ceci malgré le fait qu'il soit faux que toute sous-algébre de $\mathcal{E}^m(\mathbf{R}^n)$ fermée pour la topologie compacte d'ordre m est nécessairement fermée pour la topologie simple d'ordre m. Les sous-algèbres fermées maximales de $\mathcal{E}^m(\mathbf{R}^n)$ sont soit de la forme $\mathcal{A}_{xy} = \{f \in \mathcal{E}^m(\mathbf{R}^n);\ f(x) = f(y)\}$ pour $x, y \in \mathbf{R}^n, x \neq y$; soit de la forme $\mathcal{A}_x = \{f \in \mathcal{E}^m(\mathbf{R}^n);\ f(x) = 0\}$ pour $x \in \mathbf{R}^n$; soit de la forme

$$\mathcal{A}_{x\nu} = \left\{ f \in \mathcal{E}^m(\mathbf{R}^n);\quad \frac{\partial f}{\partial \nu}(x) = 0 \right\}$$

pour $x \in \mathbf{R}^n$ et $\nu \in \mathbf{R}^n, \nu \neq 0$ (pourvu que $m \geqslant 1$ dans ce dernier cas). Elles sont toutes de codimension 1 et leur forme est indépendante de m, si $m \geqslant 1$. Toute sous-algèbre fermée propre est contenue dans une sous-algèbre fermée maximale. Il est faux que toute sous-algèbre fermée soit l'intersection des sous-algèbres fermées de codimension finie la contenant, parce que toute sous-algèbre fermée de codimension finie est aussi fermée pour la topologie simple d'ordre m. Seules les sous-algèbres simplement fermées d'ordre m sont effectivement de telles intersections. A partir de ces résultats, $\mathcal{A}$ étant une sous-algèbre de $\mathcal{E}^m(\mathbf{R}^n)$ contenant l'unité et $\mathcal{W}$ étant un $\mathcal{A}$-sous-module de $\mathcal{E}^m(\mathbf{R}^n)$, on se demande si l'adhérence de $\mathcal{W}$ pour la topologie compacte d'ordre m est identique à son adhérence pour la topologie compacte d'ordre m seulement sur les parties compactes de $\mathbf{R}^n$ sur chacune desquelles toutes les fonctions de $\mathcal{A}$ sont des constantes. Ceci équivaut à l'assertion suivante. Si $f \in \mathcal{E}^m(\mathbf{R}^n)$ et si pour tout $\varepsilon > 0$ et toute partie compacte $K \subset \mathbf{R}^n$ sur laquelle les fonctions de $\mathcal{A}$ sont des constantes, il existe $g \in \mathcal{W}$ telle que $|D^\alpha(f-g)(x)| < \varepsilon$ quels que soient $x \in K$ et l'ordre de dérivation $\alpha = (\alpha_1, \ldots, \alpha_n)$ plus petit que m, donc $|\alpha| = \alpha_1 + \ldots + \alpha_n \leqslant m$, alors f appartient à l'adhérence de $\mathcal{W}$ dans $\mathcal{E}^m(\mathbf{R}^n)$. Bien sur, ces enoncés restent vrais pour $m=0$; et pour $m \geqslant 0$ ils contiennent comme cas particuliers le théorème sur la structure des idéaux fermés, le théorème sur la nature des sous-algèbres denses, et la forme probable du théorème analogue au théorème de Weierstrass-Stone pour $\mathcal{E}^m(\mathbf{R}^n)$. Il faut aussi analyser les résultats analogues à ceux déjà connus sur $\mathbf{R}^n$, mais cette fois ci relatifs aux fonctions réelles continuement différentiables sur des espaces différentiables plus généraux que ceux qui sont localement difféomorphes à $\mathbf{R}^n$: citons les variétés différentiables de dimension infinie (voir par exemple Lang [3]), les sous-ensembles assez généraux de $\mathbf{R}^n$ (voir Glaeser [1]). On n'a pas une étude assez complète de certaines algèbres topologiques de même nature que celles des fonctions différentiables (voir Mirkil [9], Glaeser [1], Neumark [14], Rickart [16]), soit, par exemple, celles des fonctions lipschitziennes. Quant à la forme de ces résultats en termes de synthèse spectrale d'une algèbre commutative $\mathcal{A}$ d'opérateurs continues sur un espace vectoriel topologique réel localement convexe $\mathcal{W}$, contenant

l'opérateur identité, on est assez loin de la formulation et de la démonstration de résultats généraux. Nous conjecturons le résultat suivant, contenant les théorèmes des idéaux fermés et des sous algèbres fermées sur $\mathbf{R}^n$ comme des cas particuliers. Soit $\mathcal{D}_{m+1}$ la catégorie des algèbres (sans topologies) réelles commutatives à élément unité, pures, séparées, semi-locales et d'ordre différentiel $\leqslant m+1$ (au sens de Nachbin [11]). Si $\mathcal{W}$ est localement convexe sous $\mathcal{A}$ par rapport à $\mathcal{D}_{m+1}$ et si les opérateurs de $\mathcal{A}$ sont semblablement bornés, alors, $\mathcal{V}$ étant un sous-espace vectoriel $\mathcal{A}$-invariant de $\mathcal{W}$, $\mathcal{A}'$ l'algèbre d'opérateurs induite par $\mathcal{A}$ sur l'espace localement convexe quotient $\mathcal{W}'=\mathcal{W}/\mathcal{V}$, $\mathcal{W}'$ sera localement convexe sous $\mathcal{A}'$ par rapport à la catégorie $\mathcal{D}_{m+1}$ et les opérateurs de $\mathcal{A}'$ seront semblablement bornés. En particulier, si l'algèbre topologique réelle commutative $\mathcal{A}$, à élément unité, est localement convexe par rapport à $\mathcal{D}_{m+1}$, alors toute algèbre topologique quotiente de $\mathcal{A}$ l'est aussi (dans ce cas $\mathcal{W}=\mathcal{A}$). Il peut se faire bien sûr que la validité de ces résultats de stabilité par passage au quotient reste, sous cette forme, conditionnée par des hypothèses supplémentaires, telles que celle de dimension bornée pour les espaces vectoriels des formes linéaires continues m-tangentes à $\mathcal{W}$ en chaque homomorphisme unitaire de $\mathcal{A}$ dans $\mathbf{R}$. Quoi qu'il en soit, il y a là des problèmes intéressants à élucider.

II. Théorie de Bernstein

La théorie de l'approximation de Serge Bernstein est celle de l'approximation polynomiale ponderée. Soient E un espace uniformisable séparé et N une partie fermée de E. Indiquons par $\mathcal{X}$ un ensemble filtrant de fonctions réelles positives, définies et semi-continues supérieurement sur le complémentaire $\complement N$ de N dans E. Indiquons par $\mathcal{C}\mathcal{X}_N(E)$ l'espace vectoriel des fonctions réelles continues sur E, donc sous-espace vectoriel de $\mathcal{C}(E)$, des f telles que, quels que soient la fonction $\sigma\in\mathcal{X}$ et le compact $K\subset E$, on ait $f(x)\sigma(x)\to 0$ si $x\to a$, $x\in K\cap\complement N$, pour tout point $a\in K\cap N$. Nous munirons cet espace vectoriel de la topologie définie par la famille des semi-normes

$$f\to\|f\|_{\sigma K}=\sup\{|f(x)|\,\sigma(x);\quad x\in K\cap\complement N\}.$$

On dira que $\mathcal{C}\mathcal{X}_N(E)$ est un espace vectoriel topologique ponderé de fonctions réelles continues. Remarquons que le sous-espace vectoriel de $\mathcal{C}\mathcal{X}_N(E)$, formé des fonctions nulles aux voisinages de N, est dense dans $\mathcal{C}\mathcal{X}_N(E)$. Soit $\mathcal{A}$ une sous-algèbre, contenant l'unité, de l'algèbre $\mathcal{C}(\complement N)$ des fonctions réelles continues sur l'ouvert complémentaire de N dans E; et $\mathcal{W}$ un sous-espace vectoriel de $\mathcal{C}\mathcal{X}_N(E)$ tel que l'espace vectoriel $\mathcal{W}|\complement N$, restriction de $\mathcal{W}$ à $\complement N$, soit un $\mathcal{A}$-module, ce qu'on exprimera en disant tout simplement que $\mathcal{W}$ est un $\mathcal{A}$-module. Le problème de Bernstein consiste à chercher, sous de telles conditions, une description de l'adhérence de $\mathcal{W}$ dans $\mathcal{C}\mathcal{X}_N(E)$. On appelera problème de Weierstrass le cas particulier où N est vide et $\mathcal{X}$ est réduit à la fonction constante 1; ou même, plus généralement, le cas où N est vide et $\mathcal{X}$ est arbitraire. On dira que $\mathcal{W}$ est de type fini sous $\mathcal{A}$ si l'adhérence de $\mathcal{W}$ dans $\mathcal{C}\mathcal{X}_N(E)$ coincide avec son adhérence pour la topologie sur $\mathcal{C}\mathcal{X}_N(E)$ définie par les semi-normes $f\to\|f\|_{\sigma K}$ pour $\sigma\in\mathcal{X}$ et $K\subset E$ compact tel que les fonctions de $\mathcal{A}$ soient constantes sur $K\cap\complement N$. Ceci équivaut à la condition suivante. Si $f\in\mathcal{C}\mathcal{X}_N(E)$ et si, pour tout $\varepsilon>0$, pour toute $\sigma\in\mathcal{X}$ et pour toute partie compacte $K\subset E$ telle que les fonctions de $\mathcal{A}$

soient des constantes sur $K \cap \complement N$, il existe $g \in \mathcal{W}$ telle que $|f(x) - g(x)|\sigma(x) < \varepsilon$ quel que soit $x \in K \cap \complement N$, alors f appartient à l'adhérence de $\mathcal{W}$ dans $\mathcal{C}X_N(E)$. Si N est vide, c'est à dire pour le problème de Weierstrass, $\mathcal{W}$ est toujours de type fini sous $\mathcal{A}$. On peut se poser le problème de Bernstein strict, consistant à trouver des conditions nécessaires et suffisantes pour qu'on ait la condition de type fini. Bien que le problème de Weierstrass soit complètement résolu par le théorème de Weierstrass-Stone, le problème de Bernstein général ou strict reste ouvert en dehors de quelques cas particuliers, parce qu'il donne lieu à des questions analytiques plus fines et difficiles que celles posées par le problème de Weierstrass. Ce problème strict est déjà résolu dans des situations telles que celle où E est la droite réelle $\mathbf{R}$ compactifiée par l'adjonction du point à l'infini et N est réduit au point à l'infini, $\mathcal{A}$ étant l'algèbre des polynômes réels sur $\mathbf{R}$ et $\mathcal{W}$ ayant un générateur, en tant que $\mathcal{A}$-module, réel continu partout non nul (voir Pollard [15]; voir aussi Mandelbrojt [7], Horvath [2] et Mergelyan [8] pour les questions classiques). Des conditions suffisantes générales sont connues (voir Malliavin [6], Nachbin [12], [13]). Citons le résultat suivant, lequel reflète les liens entre le problème de Bernstein dont nous nous occupons, les fonctions quasi-analytiques et le théorème de Denjoy–Carleman. Supposons que, pour toute semi-norme continue α sur $\mathcal{C}\mathcal{X}_N(E)$, $\mathcal{W}$ soit contenu dans l'adhérence dans $\mathcal{C}\mathcal{X}_N(E)$ du $\mathcal{A}$-module engendré par l'ensemble des $w \in \mathcal{W}$ tels que $\mathcal{A}$ soit contenue dans l'adhérence dans $\mathcal{C}(\complement N)$, muni de la topologie compacte, de l'algèbre engendrée par les $f \in \mathcal{A}$ vérifiant la condition

$$\sum_{m \geq 1} \frac{1}{\sqrt[m]{\alpha(f^m w)}} = +\infty.$$

Alors $\mathcal{W}$ est de type fini sous $\mathcal{A}$. On en déduit que, si $\mathcal{W}$ a un seul générateur w, en tant que $\mathcal{A}$-module, partout non nul, et si, pour ce w et pour toute semi-norme continue α sur $\mathcal{C}\mathcal{X}_N(E)$, l'ensemble des $f \in \mathcal{A}$ satisfaisant à la condition de divergence ci-dessus est séparant sur E, alors $\mathcal{W}$ est dense dans $\mathcal{C}\mathcal{X}_N(E)$. Remarquons qu'il suffit de faire toutes ces considérations dans le cas particulier où E est compact. Alors, en changeant de notation, c'est-à-dire en appelant E l'espace localement compact $\complement N$, ce qui équivaut à reléguer N à l'infini, on peut reformuler les notions et résultats précédents sur un espace localement compact E, où sont définies toutes les fonctions de $\mathcal{X}$, $\mathcal{A}$ et $\mathcal{W}$, l'ensemble N disparaisant des considérations. Terminons avec le commentaire suivant. Les deux parties de cet exposé sont des cas particuliers d'une même question, à savoir le problème de Bernstein pour les fonctions continuement différentiables. Jusqu'au présent, très peu de choses ont été écrites sur ce chapitre important de la théorie de l'approximation.

BIBLIOGRAPHIE

[1]. GLAESER, G., Étude de quelques algèbres tayloriennes. *J. Analyse Math.*, 6 (1958).

[2]. HORVATH, J., Aproximación y funciones casi-analíticas. *Publ. Fac. Ciencias*, Universidad de Madrid (1956).

[3]. LANG, S., *Introduction to Differentiable Manifolds*, 1962.

[4]. ŁOJASIEWICZ, S., Sur le problème de la division. *Studia Math.*, 18 (1959).

[5]. MALGRANGE, B., Division des distributions. *Séminaire Schwartz, Faculté des Sciences de Paris*, 4e année (1959/1960).

[6]. MALLIAVIN, P., L'approximation polynomiale ponderée sur un espace localement compact. *Amer. J. Math.*, 81 (1959).

[7]. MANDELBROJT, S., General theorems of closure. *The Rice Institute Pamphlet*, Special Issue (1951).

[8]. MERGELYAN, S. N., Weighted approximation by polynomials. *Uspehi Matematiceskih Nauk* (NS), 11 (1956); *Amer. Math. Soc. Transl.* (Ser. 2), 10 (1958).

[9]. MIRKIL, H., *The work of Šilov on Commutative Semi-simple Banach Algebras*. University of Chicago, 1951.

[10]. NACHBIN, L., Sur les algèbres denses de fonctions différentiables sur une variété, *C. R. Acad. Sci. Paris*, 228 (1949).

[11]. NACHBIN, L., Algebras of finite differential order and the operational calculus, *Ann. Math.* 70 (1959).

[12]. NACHBIN, L., On the weighted polynomial approximation in a locally compact space. *Proc. Nat. Acad. Sci. U.S.A.*, 47 (1961).

[13]. NACHBIN, L., Sur l'approximation polynomiale ponderée des fonctions réelles continues, *Atti della 2a Riunione del Groupement de Mathématiciens d'Expression Latine*, Firenze & Bologna, 1961.

[14]. NEUMARK, M. A., *Normierte Algebren*, 1959.

[15]. POLLARD, H., The Bernstein approximation problem. *Proc. Amer. Math. Soc.*, 6 (1955).

[16]. RICKART, C. E., *General Theory of Banach Algebras*, 1960.

[17]. SCHWARTZ, L., Les théorèmes de Whitney sur les fonctions différentiables. *Séminaire Bourbaki*, 3e année (1950/1951).

[18]. WHITNEY, H., On ideals of differentiable functions, *Amer. J. Math.*, 70 (1948).

THE LEVI PROBLEM IN THE THEORY OF FUNCTIONS OF SEVERAL COMPLEX VARIABLES

By RAGHAVAN NARASIMHAN

§1. The method of Oka

The history of the Levi problem is described in the lecture of H. Grauert in this Congress and I shall not repeat it. We shall rather analyze the work of Oka [6] on this problem, and see how the underlying ideas are susceptible of generalization.

Let D be an unramified domain over C^n, i.e. a connected complex manifold of dimension n together with a holomorphic map $\varphi: D \to C^n$ whose jacobian is nowhere zero. We say that D "*satisfies the Hartogs continuity condition*" if the following holds.

Every point of the relative boundary of D over C^n (which depends on φ) has a neighbourhood U such that if σ_t, $0 \leqslant t < 1$ is a continuous family of complex 1-dimensional discs in U and $\cup(\partial\sigma_t) \Subset U$, then $\cup \sigma_t \Subset U$.

Here $A \Subset X$ means that A is relatively compact in X.

A complex space X is *holomorph-convex* if for any infinite discrete set $\{x_\nu\}$, there is a holomorphic function f on x for which $f(x_\nu)$ is unbounded. If in addition holomorphic functions separate points, X is called a *Stein space*, (and we say that X is Stein).

The main theorem of Oka is then as follows.

THEOREM 1. *An unramified domain over C^n which satisfies the Hartogs continuity condition is a Stein space.*

The proof of Oka splits into several parts. To describe this proof, and for later use, we require the concept of pseudoconvex functions [6] (more often called "plurisubharmonic functions" in the literature).

Definition. An upper semicontinuous function p on an open set $\Omega \subset C^n$ is *pseudoconvex* if its restriction to any complex disc in Ω is subharmonic. It is *strongly pseudoconvex* (or just strongly convex) if to every open $\Omega' \Subset \Omega$ and any C^∞ function h on Ω, there is $\varepsilon > 0$ such that $p + \varepsilon h$ is pseudoconvex on Ω'.

We remark for later use that the definition can be extended to complex spaces by means of a local imbedding of the space into C^n.

Oka first proves the following.

LEMMA 1. *If D satisfies the continuity condition and for $x \in D$, $d(x) = d_\varphi(x)$ denotes the Euclidean radius of the largest univalent ball about x in D, then $-\log d(x)$ is pseudoconvex in D.*

Using this, he proves

LEMMA 2. *If D is as above, then there exists a strongly convex function p such that for any $\alpha > 0$, the set*

$$\{x \in D \mid p(x) < \alpha\}$$

is relatively compact in D. Moreover, p can be chosen so that locally it can be represented as the maximum of finitely many C^∞ strongly convex functions.

Using the existence of this function p on *any* domain which satisfies the continuity theorem, he proves

THEOREM 2. *Let (D,φ), $\varphi=(\varphi_1,\ldots,\varphi_n)$ be an unramified domain over C^n. Let $\delta>0$ and $D_1(\delta)=\{x\in D \mid \operatorname{Re}\varphi_1<+\delta\}$, $D_2(\delta)=\{x\in D \mid \operatorname{Re}\varphi_1>-\delta\}$. Suppose that for some $\delta>0$, $D_1(\delta)$ and $D_2(\delta)$ are Stein. Then D is Stein.*

It is a simple matter to deduce Theorem 1 from Lemma 2 and Theorem 2.

In the proof of this theorem, Oka introduced a principle which, after Grauert [3], can be formulated as follows.

A complex space such that for any $x_0\in X$, there is a holomorphic map $f:X\to C^k$ such that x_0 is an isolated point of $f^{-1}f(x_0)$ is called *K-complete.*

THEOREM 3. *Let X be a holomorph-convex, K-complete space. Then global holomorphic functions separate points and give local coordinates at every point of X.*

I would like to say that every published proof of the theorems concerning the Levi problem described in this article depend, more or less explicitly, on this principle.

§2. The Levi problem for abstract spaces

The methods of Oka make heavy use of the fact that one deals with domains over C^n. However, following an idea of Grauert [4], it is possible to generalize Oka's theorem to arbitrary complex spaces. (See [5]).

THEOREM 4. *Let X be a complex space and p a continuous strongly convex function such that for any $\alpha>0$, the set*

$$\{x\in X \mid p(x)<\alpha\}\Subset X.$$

Then X is a Stein space. On a Stein space, the function p can be chosen to be, in addition, real analytic.

For domains D on a complex space X with $D\Subset X$, it is possible to obtain a theorem where the condition on the boundary is more local [5].

Definition. Let $D\Subset X$ be an open set on the complex space X. D is called *pseudoconvex* (*strongly pseudoconvex*) if to every $x_0\in\partial D$ there is a neighbourhood U and in U a continuous pseudoconvex function (strongly convex function) p with

$$U\cap D=\{x\in U \mid p(x)<0\}.$$

THEOREM 5. *Any strongly pseudoconvex domain on a complex space is holomorph-convex and obtained from a Stein space by "blowing up" finitely many points.*

Results of the type of Theorem 4, and Oka's theorem, are established by proving first the holomorph-convexity of certain relatively compact sub-

domains defined as in Theorem 5 and then applying an approximation argument to pass to the whole space. Perhaps the most typical of these approximation theorems is the following, (see [5]).

THEOREM 6. *Let X be a Stein space and p a continuous pseudoconvex function on X. Then, for any real α, the set*

$$\{x \in X \mid p(x) < \alpha\} = X_\alpha$$

is a Stein space which is Runge in X, i.e. holomorphic functions on X_α can be approximated by holomorphic functions on X.

This theorem, when X is an unramified domain over C^n was proved by H. Behnke and K. Stein [1]. When X is a Stein manifold F. Docquier and H. Grauert [2] reduced the problem to the case of domains in C^n. These methods require one to prove first that X_α is a Stein space. The general case requires, however, a quite different method, which proves that X_α is Stein at the same time as proving that it is Runge in X [5]. I would like to add that the idea of this method is contained in the work of Oka.

The condition of *strong* pseudoconvexity in Theorems 4 and 5 is essential as is shown by an example of Grauert (see below). Very little is known about pseudoconvex domains which are not *strongly* pseudoconvex. It is perhaps possible that at least for domains with smooth boundary on a complex manifold, pseudoconvexity of the domain, together with *strong* pseudoconvexity at one point, imply its holomorph-convexity. We are however far from a proof of such a theorem.

Docquier and Grauert [2] showed that for *unramified* domains over Stein *manifolds* X, the analogue of Oka's Theorem 1 is true and their proof is in fact a reduction of the problem to Theorem 1 by imbedding X into C^n and showing that there is a neighbourhood U of X in C^n and a holomorphic map $U \to X$ which is the identity on X. This is not true even if X has one singular point, and it is not known even if pseudoconvex domains on Stein spaces are themselves Stein. The best theorem that has been proved in this direction is the following theorem, obtained in collaboration with A. Andreotti.

THEOREM 7. *If X is K-complete and p a continuous pseudoconvex function with $\{x \in X \mid p(x) < \alpha\} \Subset X$ for any $\alpha > 0$, then X is Stein.*

Another outstanding problem is the analogue of Theorem 1 for *ramified* domains over C^n.

The example given by Grauert of a pseudoconvex domain that is not holomorph-convex is as follows.

Let $n > 1$ and let Γ be the lattice in C^n generated by $w_1 = (1, 0, \ldots, 0)$, $w_j = (w_{j1}, \ldots, w_{jn})$, $j = 2, \ldots, 2n$ with $w_{j\nu} = w'_{j\nu} + i\, w''_{j\nu}$. Suppose that $w_1, \ldots, w_{2n}$ are R-independent, $w'_{j1} = 0$ for $j \geqslant 2$, and w''_{j1} $(j = 2, \ldots, 2n)$ are linearly independent over the integers.

Let T^n be the torus $C^n/\Gamma, \pi : C^n \to T^n$ the natural map. Let $U \subset C^n$ be defined by $0 < \operatorname{Re} z_1 < \frac{1}{2}$. Let

$$\varphi(z) = \frac{1}{1 - 2 \operatorname{Re} z_1} + \frac{1}{\operatorname{Re} z_1}.$$

Then φ is invariant under any $\gamma \in \Gamma$ with $(\gamma \bar{U}) \cap \bar{U} \neq \emptyset$ and so defines a function p on $D = \pi(U)$. Clearly $p \to \infty$ at the boundary of D and is pseudo-convex. But every holomorphic function f on D is constant: in fact $|f|$ has a maximum at a point x_0 on $K = \pi(\mathrm{Re}\, z_1 = \frac{1}{4})$ and there is a connected $n-1$ dimensional analytic set A through x_0 in K which is dense in K (viz. $\pi(z_1 = c_1)$, where c is such that $\pi(c) = x_0$). Hence f is constant on A and so on K. Since K has real dimension $2n-1$, f is constant on D.

REFERENCES

[1]. Behnke, H. & Stein, K., Approximation analytischer Funktionen in vorgegebenen Gebieten des Raumes von n komplexen Veränderlichen. *Nachr. Akad. Wiss. Göttingen*, 1, 15 (1939).

[2]. Docquier, F. & Grauert, H., Levisches Problem und Rungescher Satz fur Teilgebiete Steinscher Mannigfaltigkeiten. *Math. Ann.*, 140 (1960), 94–123.

[3]. Grauert, H., Charakterisierung der holomorph-vollständigen komplexen Räume. *Math. Ann.*, 129 (1955), 233–259.

[4]. —— On Levi's problem and the imbedding of real analytic manifolds. *Ann Math.*, 68 (1958), 460–472.

[5]. Narasimhan, R., The Levi problem for complex spaces. II. *Math. Ann.*, 146 (1962), 195–216.

[6]. Oka, K., Sur les fonctions analytiques des plusieurs variables. IX. Domaines finis sans points critiques intérieurs. *Japan. J. Math.*, 23 (1953), 97–155.

ОБЛАСТИ ТИПА ВЕРХНЕЙ ПОЛУПЛОСКОСТИ В ТЕОРИИ ФУНКЦИЙ МНОГИХ КОМПЛЕКСНЫХ ПЕРЕМЕННЫХ

И. И. ПЯТЕЦКИЙ-ШАПИРО

В настоящем докладе излагаются свойства некоторого класса областей, являющихся естественными аналогами обычной верхней полуплоскости. Эти области получили название областей Зигеля, в честь К. Л. Зигеля, изучившего [1], одну из наиболее замечательных областей этого класса. Первоначально области Зигеля возникли в связи с нуждами теории автоморфных функций [12]. Впоследствии же выяснилось, что наиболее эффектные приложения этих областей в теории ограниченных однородных областей. А именно оказалось, что *всякая ограниченная однородная область* в n-мерном комплексном пространстве *эквивалентна некоторой области Зигеля второго рода* (см. ниже определение) [25].

Отметим, что теория областей Зигеля по духу тесно связана с теорией бесконечномерных унитарных представлений. Например, реализация симметрических областей в виде областей Зигеля проливает свет на строение унитарных представлений и т. д. Примеров такого характера очень много.

Настоящий доклад состоит из пяти пунктов.

В пункте 1 дано определение областей Зигеля первого и второго рода. В пункте 2 — применения к теории ограниченных однородных областей. Пункт 3 посвящен алгебраическим вопросам, связанным с классификацией ограниченных однородных областей. Пункт 4 — теории спецфункций в области Зигеля. В пункте 5 — излагаются применения к теории автоморфных функций и определение областей Зигеля 3-го рода.

1. *Области Зигеля 1-го и 2-го рода.* В дальнейшем V всегда означает выпуклый конус в n-мерном вещественном пространстве, не содержащий ни одной прямой целиком. Совокупность S точек n-мерного комплексного пространства C^n вида $z=x+iy$, где $y \in V$, а x — любое называется областью Зигеля 1-го рода. Простейшим примером служит обычная верхняя полуплоскость.

Простейшим примером области Зигеля 2-го рода является область: $\operatorname{Im} z-|u|^2>0$, где z,u — числовые комплексные переменные. Последняя область аналитически эквивалентна шару $|z_1|^2+|z_2|^2<1$ и весьма удобна во всех вопросах, связанных с изучением геометрии шара вблизи фиксированной точки границы. Прежде чем давать общее определение областей Зигеля 2-го рода, мы введем понятие V-эрмитовых вектор-функций $F(u,v)$. Функция $F(u,v)$ от пары векторов $u,v \in C^m$ со значениями в C^n называется V-эрмитовой, если: 1) $F(u,v)$ — линейная функция от u, 2) $F(u,v)=\overline{F(v,u)}$ (эрмитовость), 3) $F(u,u) \in \overline{V}$ (1) («положительность»), 4) $F(u,u)=0$ только при $u=0$.

(1) $\overline{V}$ — замыкание конуса V.

Областью Зигеля 2-го рода называется совокупность S всех точек $(z,u) \in C^{n+m}$, для которых $\operatorname{Im} z - F(u,u) \in V$ где V — заданный конус, а F — заданная V-эрмитова форма.

Приведем простейший пример. В качестве V возьмем конус в трехмерном пространстве, который задается неравенствами: $y_1 y_2 - y_3^2 > 0$, $y_1 > 0$. $F(u,v)$ определим следующими формулами. Обозначим через (u,v) — эрмитово скалярное произведение двух векторов в C^m. Положим $F_1(u,v) = (u,v)$, $F_2(u,v) = F_3(u,v) = 0$. В этом примере $n=3$, m — любое. Можно показать, что все эти области аналитически однородны, не симметрические (см. п. 2) и аналитически эквивалентны ограниченным областям. Последний факт легко обобщается. А именно, можно доказать, что *всякая область Зигеля 1-го и 2-го рода* аналитически *эквивалентна* некоторой ограниченной области. Разумеется, не всякая область Зигеля аналитически однородна. Повидимому, можно показать, что аналитически однородная область Зигеля однородна также относительно линейных преобразований.

Применение областей Зигеля связано так или иначе со сравнительной простотой описания в них полной группы линейных преобразований. В каждой области Зигеля действуют следующие аналитические преобразования

$$z \to z + a + 2\,i\,F(u,b) + i\,F(b,b) \quad (u \to u + b), \tag{1}$$

где a — n-мерный действительный вектор, b — m-мерный комплексный вектор. Такие преобразования принято называть «параллельными переносами».

Пусть $y \to Ay$ и $u \to Bu$ — линейные преобразования со следующими свойствами: 1) A — сохраняет конус V, 2) $F(Bu,Bv) = A(F(u,v))$. Легко видеть, что преобразование

$$z \to Az, \quad u \to Bu \tag{2}$$

переводит область Зигеля в себя. Верно и обратно, а именно всякое линейное преобразование, переводящее область Зигеля в себя, является композицией преобразований вида (1) и (2). Эта простая теорема имеет многочисленные приложения.

Отметим следующую нерешенную проблему. *Когда две области Зигеля аналитически эквивалентны между собой?* Мы предполагаем, что это имеет место только в случае, когда они линейно эквивалентны. Легко показать, что последнее означает, что их конусы и вектор-функции линейно эквивалентны.

В заключение отметим, что в областях Зигеля сравнительно просто и эффективно пишутся интегральные формулы и, в частности, ядро Бергмана. Наиболее законченные результаты имеют место для аффинно однородных областей Зигеля. Они принадлежат С. Г. Гиндикину [19] и [20]. Отметим, что ряд результатов Хуа Ло Кена [18] о симметрических областях может быть значительно проще получен, если использовать реализацию их в виде области Зигеля 2-го рода.

2. *Ограниченные однородные области.* В 1935 г. Э. Картан [2] нашел все ограниченные однородные области в C^2 и C^3. Оказалось, что все они

являются симметрическими(1) и, следовательно, полные группы аналитических автоморфизмов этих областей — полупростые группы Ли.

Э. Картану удалось дать полную классификацию ограниченных симметрических областей в комплексном пространстве любой размерности. Им была поставлена следующая проблема: существуют ли ограниченные однородные несимметричные области.

А. Борель [3] показал, что если в ограниченной однородной области транзитивно действует полупростая группа Ли, то эта область симметрическая. Ж. Козюль [4] получал из других соображений тот же результат. Эти результаты были несколько усилены Хано [6].

В 1959 г. автором настоящего доклада был построен пример несимметрической области в четырехмерном комплексном пространстве [13]. Такой областью оказалась область, рассмотренная в пункте 1 с $m=1$. Впоследствии удалось показать [17], что всякая ограниченная однородная область, в которой транзитивно действует разрешимая группа Ли, аналитически эквивалентна аффинно-однородной области Зигеля 1-го или 2-го рода. Одновременно было установлено, что аффинно-однородные области Зигеля аналитически эквивалентны тогда и только тогда, когда они аффинно-эквивалентны. Сравнительно недавно С. Г. Гиндикину, Э. Б. Винбергу и автору настоящего доклада удалось получить окончательный результат, а именно — всякая ограниченная однородная область аналитически эквивалентна некоторой аффинно-однородной области Зигеля 1-го или 2-го рода [25]. Из этой теоремы следует, что всякая ограниченная однородная область гомеоморфна эвклидову пространству. Из этой теоремы также следует, что всякая ограниченная однородная область аналитически эквивалентна некоторой аффинно-однородной области. Интересные результаты об аффинно-однородных областях получил недавно Козюль [5].

3. *Классификация ограниченных однородных областей.* Классификация ограниченных симметрических областей, принадлежащая Э. Картану, содержит полное перечисление последних. Для всех ограниченных однородных областей положение сложней: не удается составить таблицу, в которой бы все они содержались. Тем не менее построенная в настоящее время теория может считаться заменяющей классификацию в виде таблицы, т. е. она позволяет выяснить различные конкретные вопросы об областях, например, строение полной группы автоморфизмов, вид интегральных формул, дифференциально-геометрические свойства и т. д.

Основой классификацией является следующая теорема. *В каждой ограниченной однородной области просто транзитивно действует абсолютно вещественная*(2) *группа Ли* $\mathfrak{G}$. Группа $\mathfrak{G}$ определена с точ-

---

(1) Область D называется симметрической, если для любой точки $z_0 \in D$ существует такое аналитическое взаимно однозначное отображение φ_0 области D на себя со следующими свойствами:

1) $\varphi_0(z)=z$ тогда и только тогда, когда $z=z_0$,

2) $\varphi_0^2(z)=z$ при всех z.

(2) Вещественная группа Ли называется абсолютно-вещественной, если у всех преобразований в присоединенном представлении все собственные значения вещественны. Легко показать, что абсолютно вещественная группа Ли разрешима.

ностью до замены на сопряженную *однозначно*. Пользуясь этой теоремой можно установить взаимнооднозначное соответствие между ограниченными однородными областями и вещественными алгебрами Ли с некоторыми дополнительными свойствами, получившими название j-алгебр. Приведем определение j-алгебр. Пусть G — векторное пространство над полем вещественных чисел. Предположим, что в G определено строение алгебры Ли (с операцией коммутирования $x,y\to[x,y]$) и строение комплексного эрмитова пространства (с эндоморфизмом j и эрмитовым скалярным произведением $h(x,y)$). Предположим также, что выполняются соотношения:

1) $[x,y]+j[jx,y]+j[x,jy]-[jx,jy]=0$ при любых $x,y\in G$
2) существует на G такая линейная форма ω, что
$$\operatorname{Im} h(x,y)=\omega([x,y])$$
3) для любого $x\in G$ собственные значения преобразования $y\to[x,y]$ все вещественны.

Алгебра Ли G, обладающая всеми указанными выше свойствами, называется j-алгеброй(1). Изучение j-алгебр по существу основано на исследовании свойств системы корней. А именно, каждую j-алгебру G можно представить в виде суммы $H+K$, где K — коммутант, а H — дополнительная подалгебра. K — совпадает с максимальным нильпотентным идеалом, а H — коммутативна, причем все элементы из H — полупросты. Пространство K разлагается в сумму корневых пространств.

$[h,x]=\alpha(h)x$ при всех $x\in K_\alpha$ линейная форма $\alpha(h)$ называется корнем.

Можно показать, что системы корней всегда имеют следующий вид:

α_m, $\frac{1}{2}\alpha_m$, $\frac{1}{2}(\alpha_k-\alpha_m)$, $\frac{1}{2}(\alpha_k+\alpha_m)$ $(k<m, k,m=1,\dots n)$ где $\alpha_1,\dots,\alpha_n$ — линейно независимые векторы.

Кратности N_α корневых пространств, вообще говоря, отличны от единицы. Между ними есть соотношения. Например, $N_{\frac{1}{2}(\alpha_k-\alpha_m)}=N_{\frac{1}{2}(\alpha_k+\alpha_m)}$, $N_{\frac{1}{2}\alpha_k}$ — четно, $N_{\alpha_k}=1$. Кроме того, справедливы некоторые неравенства. Система инвариантов — N_α, как правило, не определяет полностью j-алгебру, за исключением специальных случаев. Дополнительные инварианты более сложны, однако, также поддаются довольно детальному изучению.

Для описания j-алгебр и соответствующих им областей также удобен аппарат обобщенных матричных алгебр, предложенный Э. Б. Винбергом. Обобщенной матричной алгеброй называется биградуированная алгебра с инволюцией. Вообще говоря, эти алгебры не ассоциативны.

Э. Б. Винбергом был введен специальный класс обобщенных матричных алгебр — T-алгебры. Им было показано, что всякий выпуклый однородный конус можно рассматривать как совокупность всех «положительно определенных эрмитовых» матриц некоторой T-алгебры [24]. Это соответствие между конусами и T-алгебрами взаимно-однозначно. Матричное исчисление Э. Б. Винберга тесно связано с разбиением алгебры G на корневые подпространства. Именно, матричные элементы являются элементами соответствующим образом упорядоченных корневых подпространств, а умножение определяется с помощью операции коммутирования и эндоморфизма j.

---

(1) В [14] и [15] — j-алгебрами назван более общий объект.

4. Специальные функции. Излагаемые в этом пункте результаты принадлежат в основном С. Г. Гиндикину [20]. Они являются обобщением и развитием некоторых результатов, полученных для различных частных случаев в работах [11], [18].

Пусть V выпуклый аффинно-однородный конус. $\mathfrak{G}$ — просто транзитивная абсолютно вещественная группа Ли. Обобщенной степенной функцией $f(y)$, $y \in V$ условимся называть такую функцию, что $f(gy) = \chi(g)f(y)$, где $\chi(g)$ — некоторый характер группы $\mathfrak{G}$. Для дальнейшего удобно записывать обобщенную степенную функцию в виде y^ϱ, где ϱ — означает гомоморфизм алгебры Ли G группы $\mathfrak{G}$ в поле комплексных чисел. Совокупность всех гомоморфизмов, очевидно, образует n-мерное комплексное пространство. Функция y^ϱ в дальнейшем нормирована условием, что в некоторой фиксированной точке y_0 ее значение равно 1. Вводятся и исследуются следующие функции:

1) Гамма-функция

$$\Gamma_V(\varrho) = \int_V \exp(-l(y))\, y^\varrho\, dy,$$

где $l(y)$ — некоторая линейная функция на V, а dy здесь и в дальнейшем инвариантная мера на V.

2) Бэта-функция

$$\mathrm{B}_V(\varrho, \sigma; a) = \int_{V_a} y^\varrho (a-y)^{\sigma+d}\, dy,$$

$$\mathrm{B}_V(\varrho, \sigma) = \mathrm{B}_V(\varrho, \sigma; y_0)$$

где $a \in V$, а V_a — означает совокупность точек $y \in V$ таких, что $a - y \in V$. V_a — естественный аналог отрезка $[0, a]$, $y^{-d}dy$ — эвклидова инвариантная мера.

3) Гипергеометрическая функция

$$F_V(\varrho, \sigma, \tau; a) = \frac{1}{\mathrm{B}_V(\sigma, \tau-\sigma; a)} \int_{V_a} y^\sigma (y_0 - y)^{-\varrho} (a-y)^{\tau-\sigma+d}\, dy.$$

Для этих функций имеют место аналоги известных классических соотношений. Например,

$$\mathrm{B}_V(\varrho, \sigma) = \frac{\Gamma_V(\varrho)\,\Gamma_V(\sigma)}{\Gamma_V(\varrho+\sigma)}.$$

Нам кажется, что эта теория будет иметь многочисленные приложения.

5. Применения к теории автоморфных функций. В настоящем пункте дано определение областей Зигеля 3-го рода и применения этого понятия к теории автоморфных функций. Последние основаны на следующей теореме. Пусть D — ограниченная симметрическая область. Для любой точки z_0 — границы области D существует аналитическое отображение D на некоторую неограниченную область S (область Зигеля

3-го рода), обладающее следующим свойством: все аналитические автоморфизмы области D, оставляющие на месте данную точку z_0 переходят в линейные преобразования области D.

Перейдем теперь к определению областей Зигеля 3-го рода. Отметим, что это понятие более сложно и менее устоялось, чем понятие областей Зигеля 1-го и 2-го рода. Пусть H — область Зигеля 2-го рода, заданная как обычно, соотношением

$$\operatorname{Im} z - F(u,u) \in V, \quad z \in C^n, \quad u \in C^m \tag{1}$$

Обозначим через $\overline{K}$ совокупность всех комплексно линейных преобразований k пространства C^m со следующими свойствами:

$$F(k\bar{u}, v) = F(k\bar{v}, u) \quad \text{при всех} \quad u, v \in C^m \tag{2}$$

$$F(u,u) - F(k\bar{u}, k\bar{v}) \in \overline{V} \quad \text{при любом} \quad u \in C^m. \tag{3}$$

Пусть далее K — означает совокупность всех внутренних точек $\overline{K}$.

Пусть D — некоторая ограниченная область в C^k, а $t \to k(t)$ — аналитическое отображение D в K.

Областью Зигеля 3-го рода с базой D и слоем H называется совокупность (z,u,t) точек $C^n \times C^m \times C^k$ для которых

$$\operatorname{Im} z - \operatorname{Re} F(u, (1 - k(t)\overline{k(t)})^{-1}(u - k(t)\bar{u})) \in V \tag{4}$$

$$t \in D. \tag{5}$$

Легко показать, что если $k \in K$, то преобразование $1 - k\bar{k}$ невырождено.

Таким образом, чтобы задать область Зигеля 3-го рода надо задать конус V, вектор-функцию F, область D и аналитическое отображение (вообще говоря, невзаимооднозначное) области D в K.

Например, если $F(u,v) = (u,v)$ — обычное эрмитово произведение в C^m, то K — означает совокупность всех комплексных симметрических матриц k порядка m, таких что матрица $1 - k\bar{k}$ — положительно определена, т. е. так называемый круг Зигеля.

Простейшим примером области Зигеля 3-го рода служит следующая область

$$\operatorname{Im} z - (1 - |t|^2)^{-1}(|u|^2 + \operatorname{Re} \bar{t}u^2) > 0,$$

$$|t| < 1,$$

где z, u, t — числовые комплексные переменные.

В областях Зигеля 3-го рода вместо линейных преобразований естественно рассматривать более широкий класс «квазилинейные преобразования». Так называются преобразования вида:

$$z \to A(t)z + a(u,t), \quad u \to B(t)u + b(t), \quad t \to g(t), \tag{6}$$

где $A(t), B(t)$ аналитические в D матричные функции, $a(u,t)$ и $b(u,t)$ — аналитические вектор-функции, $t \to g(t)$ — аналитический автоморфизм области D. Можно указать, аналогично тому как это было сделано

нами в пункте 1 для областей 1-го и 2-го рода условия аналитической однородности [см. 16]. Мы не будем здесь на этом останавливаться. Отметим, что аналитически эквивалентные области Зигеля 3-го рода, вообще говоря, «различны», т. е. не отображаются друг на друга «квазилинейно». Можно показать, однако, что для всякой ограниченной однородной области существует конечное число реализаций в виде области Зигеля 3-го рода и все их перечислить. Это сводится к перечислению всех гомоморфизмов алгебры Ли, определенной в пункте 3, перестановочных с j.

Сформулированная вначале этого пункта теорема, вытекает из следующего более сильного результата. Пусть D — ограниченная симметрическая область, F — компонента границы, т. е. максимальный аналитический кусочек границы области D. Всегда существует область Зигеля 3-го рода S с базой F, аналитически эквивалентная D, причем все аналитические автоморфизмы D, переводящие F в себя, являются «квазилинейными» преобразованиями S. Эта теорема позволяет изучать поведение автоморфных функций вблизи границы [12], [16].

ЛИТЕРАТУРА

[1]. Siegel, C. L., Symplectic Geometry. *Amer. J. Math.*, 65 (1943), 1–86.

[2]. Cartan, E., Sur les domaines bornés homogènes de l'espace de n variables complexes. *Abh. Math. Sem. Hamburg*, 11 (1935), 116–162.

[3]. Borel, A., Kählerian coset spaces of semisimple Lie groups. *Proc. Nat. Acad. Sci. U.S.A.*, 40 (1954), 1147–1151.

[4]. Koszul, I. L., Sur la forme hermitienne canonique des espaces homogènes complexes. *Can. J. Math.*, VII, 4 (1955).

[5]. —— Domaines bornés homogènes et orbites de groupes de transformations affines. *Bull. Soc. Math. France.*, 89 (1961), 515.

[6]. Hano, I., On Kählerian homogeneous spaces of unimodular Lie groups. *Amer. J. Math.*, 79, 4 (1957), 885–900.

[7]. Koecher, M., Positivitätsbereiche im R^n. *Amer. J. Math.*, 79 (1957), 575–596.

[8]. —— Die Geodätsbereiche von Positivitätsbereichen. *Math. Ann.*, 135 (1958), 192–202.

[9]. —— Analysis in reellen Jordan-Algebren. *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* (1958), 67–74.

[10]. Rothaus, O., Domains of positivity. *Abh. Math. Sem. Hamburg*, 24 (1960), 189–235.

[11]. Herz, C. S., Bessel functions of matrix argument. *Ann. Math.*, 61: 3 (1955), 474–523.

[12]. Пятецкий-Шапиро, И. И., Об оценке размерности пространства автоморфных форм для некоторых типов дискретных групп. *ДАН СССР*, 113, 5 (1957), 980–983.

[13]. —— Об одной проблеме Э. Картана. *ДАН СССР* 124, 2 (1959), 272–273.

[14]. —— О ограниченных однородных областях. I. *Изв. АН СССР. Сер. мат.*, 26 (1962) 107–124.

[15]. —— Строение j-алгебр *Изв. АН СССР. Сер. мат.*, 26 (1962) 453–484.

[16]. —— *Геометрия классических областей и теория автоморфных функций.* М. 1961.

[17]. —— О классификации ограниченных однородных областей в n-мерном комплексном пространстве. *ДАН СССР*, 141, 2 (1961), 316–319.

[18]. Хуа-Ло-Кен, Гармонический анализ функций многих комплексных переменных в классических областях. М. и Л. 1959.

[19]. Гиндикин, С. Г., Интегральные формулы для областей Зигеля второго рода. *ДАН СССР*, 141, 2 (1961), 531–534.

[20]. —— Интегральные формулы в комплексных однородных ограниченных областях. *УМН*, XVII, 3 (1962), 209–211.

[21]. Винберг, Э. Б., Однородные конусы. *ДАН СССР*, 133, 1 (1960), 9–12.

[22]. —— Теорема Морозова–Бореля для вещественных групп Ли. *ДАН СССР*, 141, 2 (1961), 270–273.

[23]. —— Выпуклые однородные области. *ДАН СССР*, 141, 3 (1961), 521–524.

[24]. —— Автоморфизмы однородных выпуклых конусов. *ДАН СССР*, 143, 2 (1962).

[25]. Винберг, Э. Б., Гиндикин, С. Г., и Пятецкий-Шапиро, И. И., О классификации и канонической реализации однородных ограниченных областей в комплексном пространстве. Труды Московского математического общества. 12 (1963).

UNIQUE CONTINUATION THEOREMS FOR SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS

By A. PLIŚ

In this lecture I intend to present essential ideas and not the complete or detailed formulations of theorems. Individual results will be cited for illustration and many important papers will be not discussed especially if they were later generalized.

1. Unique continuation property over surfaces

To start with I recall the definition of the unique continuation property. A family of functions of n variables has the property of unique continuation over a surface S if (roughly speaking) any two functions of this family equal on one side of S are equal on a neighbourhood of S. To make the definition precise let the surface S cut an open set Z into two open sets A, B (i.e. $Z-S=A \cup B$, $A \cap B=\emptyset$). We say that the family F of functions defined on open subsets of Z possesses the property of unique continuation over the surface S from A to B at a point P in S if any two functions in F defined on a neighbourhood N of P and equal on $A \cap N$ are equal on a neighbourhood of P.

Given a partial differential equation or a system of such equations and a surface S the problem arises whether the family of all solutions (or a certain subfamily of this family) possesses this property. The answer to this problem depends on both the equation and the surface. If the property is satisfied we shall say that the equation possesses (or the solutions possess) property U over surface S from A to B (at the point P).

For illustration let us consider a trivial partial differential equation

$$\frac{\partial u}{\partial t}=0 \tag{1}$$

on the plane t,x. Each function of the form $u(t,x)=f(x)$ satisfies (1) and therefore property U is not fulfilled over lines parallel to t-axis (which are characteristics for (1)). However, equation (1) possesses property U over any straight line not parallel to t-axis.

Under the additional assumption that for arbitrary data on a sufficiently smooth non-characteristic surface S there exists a solution of Cauchy's problem on a neighbourhood of S, property U is equivalent to the uniqueness of solutions for a suitable Cauchy problem with data on S. Obviously for linear homogeneous partial differential equations property U is equivalent without additional assumptions to the local vanishing of a solution for the Cauchy problem with zero data on S.

In some unique continuation theorems the surface is supposed not only to be non-characteristic but also to be convex, (i.e. the set B is supposed to be convex). In certain cases it is not known whether the convexity assumption is essential.

Now we shall deal exclusively with linear equations leaving remarks on non-linear equations to the very end of this lecture.

Sixty years ago Holmgren [4] proved that the solutions of class C^m for linear partial differential equations of order m with analytic coefficients possess property U over non-characteristic surfaces.

For the case of non-analytic coefficients the problem is so far only partly solved. In this case unique continuation theorems over non-characteristic surfaces hold for certain classes of equations provided characteristics are simple. For systems of equations with leading coefficients real and first derivatives of coefficients Hölder-continuous the result was obtained by Calderón [2] and for elliptic equations and Lipschitz-continuous coefficients by Hörmander [6].

For elliptic differential equations the unique continuation theorems can be considered as generalization of the well known fact that analytic functions of the complex variable $z, f(z)=f(t+ix), tx$ real, which are solutions of the (Cauchy–Riemann) elliptic equation

$$\frac{\partial f}{\partial t}+i\frac{\partial f}{\partial x}=0$$

have unique continuation property over any surface.

Even a partial differential equation of the first order

$$\frac{\partial z}{\partial t}+a(t,x)\frac{\partial z}{\partial x}=0$$

with the complex coefficient $a(t,x)$ of class C^∞ need not possess property U over non-characteristic surface S if it is not of elliptic type on a neighbourhood of the surface S. Evidently such a complex equation has simple characteristics, but the equivalent system of two real partial differential equations (for two unknown functions $\mathrm{Re}(z(t,x))$ and $\mathrm{Im}(z(t,x))$, where $\mathrm{Re}(z)$, $\mathrm{Im}(z)$ denote real and imaginary part of z respectively) of the first order has double characteristics on S.

The elliptic differential equation with at most double characteristics may also lack property U. An equation without property U of the form

$$\left[\left(\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)^2+t\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)^2-\frac{1}{2}\frac{\partial^4}{\partial x^4}\right]u=f\frac{\partial u}{\partial x}+gu,$$

the surface S being the plane $t=0$, can be constructed [10]. The equation is real, the coefficients f,g are of class C^∞ on R^3. Only the coefficients of the first and the zero order derivatives are non-analytic.

The Lipschitz-continuity in Hörmander's result can not be replaced by Hölder-continuity even for the real elliptic differential equation of the second order. This may be shown by an example of the form

$$\frac{\partial^2 u}{\partial t^2}+\frac{\partial^2 u}{\partial x^2}+\mathbb{1}(t)\frac{\partial^2 u}{\partial y^2}=f\frac{\partial u}{\partial x}+g\frac{\partial u}{\partial y}+hu,$$

where $l(t)$ is positive, Hölder-continuous and independent of x,y, the coefficients f,g,h being real and of class C^∞ on R^3 [11].

For elliptic differential equations with constant leading coefficients property U was proved by Hörmander [5] also in the case of at most double characteristics.

The known examples of elliptic differential equations in two independent variables without property U are not as satisfactory as those in three independent variables. The relevant equation of class C^∞ has fourfold characteristics. Specifically it has the form

$$\left[\left(\frac{\partial}{\partial t}-i\frac{\partial}{\partial x}\right)^4+t^4\left(\frac{\partial}{\partial x}\right)^4+i\left(\frac{\partial}{\partial x}\right)^3\right]u=f(t,x)u,$$

where $f(t,x)$ is a complex function of class C^∞ [10].

I emphasize that only this coefficient is non-analytic. It shows that in Holmgren's theorem the assumption that coefficients are analytic is essential even for a coefficient of zero order derivative.

Examples with smaller multiplicity but not of class C^∞ were constructed independently by Cohen and me [9]. Douglis [3] showed that property U holds for elliptic equations in two independent variables with at most double characteristics and coefficients of class C^1 under the additional assumption that the equation can be transformed to a Lipschitz-continuous canonical form by a mapping of class C^1. The theorem does not hold without additional assumption but it is not known whether this assumption can be replaced by the assumption that the coefficients are of class C^2. Similar results for arbitrary number of independent variables were obtained by Calderón.

2. Conjecture

Consider a linear elliptic partial differential equation

$$P\left(t,x,\frac{\partial}{\partial t},\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)u(x)=0,\ x=(x_1,\ldots,x_n) \tag{2}$$

of order m with coefficients continuous on a neighbourhood N of the origin.

We shall now formulate a condition for characteristic roots of (2) considered as functions of t. Let $z_j(t,x,q), j=1,\ldots,m$, be continuous complex functions of t such that given $t,x,q,(t,x)\in N, q$ an arbitrary complex vector $q=(q_1,\ldots,q_n)$, the numbers $z_j=z_j(t,x,q), j=1,\ldots,m$, constitute the complete set of solutions of the algebraic equation $P(t,x,z,q)=0$ for the unknown z. Consider the following system of inequalities:

$$|\operatorname{Re}(z_j(t,x,q)-z_j(s,x,q))|\leqslant K\max\left(|t-s||q|,|q|^{1/2}\right), \tag{3}$$

$$j=1,\ldots,m,\quad \text{for}\quad |q|\geqslant K,\ \frac{|\operatorname{Re}(q)|}{|q|}\leqslant\varepsilon,\ (t,x),(s,x)\in N,$$

where K,ε are certain positive constants.

In the known examples of elliptic equations not possessing property U over the plane $t=0$ inequalities (3) are not satisfied. This fact as well as a

certain method of proof for unique continuation theorems suggest that a theorem of the following type may be true.

Conjecture C. An elliptic differential equation with coefficients continuous on a neighbourhood of the origin and satisfying condition (3) possesses property U over the plane $t=0$ (or over the surface $t=|x|^2$ from the set $t<|x|^2$ to the set $t>|x|^2$) at the origin.

If conjecture C is valid many known and new unique continuation theorems formulated with the notion of multiplicity of characteristics follow. The theorems for equations with simple characteristics and for equations with at most double characteristics and constant leading coefficients mentioned above follow from conjecture C. Another corollary of conjecture C is the unique continuation theorem for operators which are products of two elliptic operators with simple characteristics obtained by Mizohata [7]. Instead of two we may take the product of any number of operators with simple characteristics.

A unique continuation theorem for linear elliptic differential equations with at most triple characteristics, leading coefficients constant and other coefficients of class C^1 results from conjecture C too. The problem, whether the latter theorem holds true, is open.

In conjecture C condition (3) replaces assumptions on the multiplicity of characteristics.

A modified conjecture is obtained if condition (3) is strengthened to the following Lipschitz-type condition

$$|\mathrm{Re}(z_j(t,x,q)-z_j(s,x,q))|\leqslant K|t-s||q| \quad (j=1,\dots,m) \tag{4}$$

But the full strength of conjecture C is needed to obtain corollaries in the general case of non-simple characteristics. For instance Hörmander's theorem for equations with constant leading coefficients and at most double characteristics does not follow from the variant of conjecture C, in which inequalities (3) are replaced by inequalities (4).

One can also replace condition (3) by a weaker one and obtain a one-sided unique continuation conjecture i.e. an hypothesis implying property U over planes $t=c$ from half-space $t\leqslant c$ to half-space $t\geqslant c$ but not from half-space $t\geqslant c$ to half-space $t\leqslant c$.

A result similar to that conjectured as C with a variant of inequalities (3) in which the constant K is replaced by the term of the form $\delta t^{-1}, \delta$ being a positive constant, $\delta<1$ might yield stronger results in the case of two independent variables.

Obviously condition (3) and its variants can be considered for elliptic systems of partial differential equations. It can be also reformulated for non-elliptic equations, for instance in the case that multiple characteristics are non-real.

3. Results of slightly different character

The uniqueness in Cauchy's problem is easier to establish for certain subclasses of solutions for instance in the case of linear homogeneous equations and zero data in the class of non-negative solutions. I cite here one theorem [8]. If the coefficients of a system of partial differential equations of the first order are Lipschitz-continuous with respect to suitable variables

and the components of a solution are non-negative then they vanish on a neighbourhood of a non-characteristic surface S if they vanish on S. The Lipschitz-continuity is essential here. No assumptions on the type of the system or multiplicity of characteristics are made in the theorem.

Properties analogous to property U have also been studied. For the linear homogeneous second order equation of elliptic type with Lipschitz-continuous coefficients the strong unique continuation theorem was proved by Aronszajn, Krzywicki and Szarski [1], i.e. they proved that if a solution of such an equation has a zero of infinite order at an interior point of its domain of definition it vanishes identically. This kind of unique continuation property is evidently stronger than property U. It is not known whether this second property holds for non-analytic linear elliptic differential equations of higher order in the case of simple characteristics or more generally in the case when a condition of type (3) is satisfied. For non-elliptic equations the strong unique continuation property need not to be satisfied. As an example even the trivial partial differential equation $\partial u(t,x)/\partial t=0$ may be taken, and its solution $u(t,x)=f(x)$, where $f(x)$ has zero of infinite order at the point $x=0$ and $f(x)\neq 0$ for $x\neq 0$.

Now we pass over to discuss the following property W weaker than property U: any solution which vanishes outside a certain compact set vanishes identically. Property W need not be satisfied even for real elliptic partial differential equations of class C^∞ [10] but we have the following well known theorem: For any elliptic linear homogeneous partial differential equation continuous on a neighbourhood of the origin there exists a neighbourhood N of the origin such that every solution of the equation vanishing outside N vanishes identically. The last theorem fails for non-elliptic partial differential equations [12].

The unique continuation problems for non-linear partial differential equations can be reduced to linear problems in a known way but the problems for solutions which are not sufficiently smooth may need separate treatment because the smoothness of coefficients of the linear equations to be considered depends not only on the smoothness of the non-linear equations but also on the smoothness of the solutions. This effect is especially important for non-elliptic equations, the solutions of smooth non-elliptic equations not necessarily being smooth.

REFERENCES

[1]. Aronszajn, N., Krzywicki, A. & Szarski, J., A unique continuation theorem for exterior differential forms on Riemannian manifolds. *Arkiv för Matematik*, 4 (1962), 417–453.

[2]. Calderón, A. P., Uniqueness in the Cauchy problem for partial differential equations. *Amer. J. Math.*, 80 (1958), 16–36.

[3]. Douglis, A., On uniqueness in Cauchy problems for elliptic systems of equations. *Comm. Pure Appl. Math.*, 13 (1960), 593–607.

[4]. Holmgren, E., Über Systeme von linearen partiellen Differentialgleichungen. *Öfv. K. Vetensk. Akad. Förh.* 58 (1901), 91–103.

[5]. Hörmander, L., On the uniqueness of the Cauchy problem. *Math Scand.*, 6 (1958), 213–225.

[6]. —— On the uniqueness of the Cauchy problem. II. *Math. Scand.*, 7 (1959), 177–190.

[7]. MIZOHATA, S., Unicité du prolongement des solutions des équations elliptiques du quatrième ordre. *Proc. Japan Acad.*, 34 (1958), 687–692.

[8]. PLIŚ, A., On the uniqueness of the non-negative solution of the homogeneous Cauchy problem for a system of partial differential equations. *Ann. Polon. Math.*, 2 (1955), 314–318.

[9]. —— Non-uniqueness in Cauchy's problem for differential equations of elliptic type. *J. Math. Mech.*, 9 (1960), 557–562.

[10]. —— A smooth linear elliptic differential equation without any solution in a sphere. *Comm. Pure Appl. Math.*, 14 (1961), 599–617.

[11]. —— On non-uniqueness in Cauchy problem for an elliptic differential equation of the second order. (To appear in *Bull. Acad. Polon. Sci.*)

[12]. —— Homogeneous partial differential equations possessing solutions with arbitrary small supports. (To appear in *Bull. Acad. Polon. Sci.*)

CORRECTLY POSED PROBLEMS FOR LINEAR CONSTANT COEFFICIENT PARTIAL DIFFERENTIAL EQUATIONS IN THE REAL HALF-SPACE

By G. E. SHILOV

There are many different correctly posed classical boundary problems for partial differential equations, e.g. the Cauchy problem for the wave equation or the heat equation, and the Dirichlet problem for the Laplace equation. However, for some equations such as the ultrahyperbolic one, no correctly posed boundary problem in the half-space was known until recently (in the non-analytic case).

In [1] we have given a correctly posed boundary problem for every differential equation

$$\frac{\partial^m u(x,t)}{\partial t^m} = \sum_{k=0}^{m-1} P_k\left(i\frac{\partial}{\partial x}\right)\frac{\partial^k u(x,t)}{\partial t^k} \tag{1}$$

with constant coefficients ($x=(x_1,\ldots,x_n)$). This problem is stated in the Bochner class H of generalized functions f in R^n which belong to L^2 and all their formal derivatives. The Fourier image $\tilde{H}=\mathfrak{F}H$ of H is the class of functions $g\in L^2$ which remain in L^2 and their products by any polynomial. Let Q_k be the set of all $\sigma\in R^n$ such that the equation

$$\lambda^m = \sum_{k=0}^{m-1} P_k(\sigma)\lambda^k \tag{2}$$

has at least k roots with $\operatorname{Re}\lambda\leqslant 0$, and denote by $H(Q_k)$ the subset of $L^2(Q_k)$ formed by restrictions of functions in H. These spaces as well as H have a natural topology. The boundary problem of [1] is to find a function $u(t,x)$, $0\leqslant t<\infty$, $x\in R^n$, such that

(*a*) u is m times continuously differentiable as a function of t with values in H and does not grow faster than a power of t when $t\to\infty$;

(*b*) u satisfies (1) and the boundary conditions

$$\mathfrak{F}u_k = g_k \quad \text{in} \quad Q_k \quad (k=1,\ldots,m), \tag{3}$$

where g_k are given functions in $\tilde{H}(Q_k)$ and $u_k(x)=\partial^{k-1}u(0,x)/\partial t^{k-1}$. It was proved in [1] that a unique solution exists and that it depends continuously on the initial data.

The boundary problem can be reformulated in terms of the functions $u(0,x),\ldots,\partial^{m-1}u(0,x)/\partial t^{m-1}$ instead of their Fourier transforms if and only if the number r of roots of (2) with $\operatorname{Re}\lambda\leqslant 0$ is independent of σ; (3) then means simply that $\partial^k u(0,x)/\partial t^k$ is given in H for $k<r$. If $r=m$ the equation (1) was studied by Petrowsky, if $r=m/2$ it is elliptic.

A solution with initial data $u(0,x) = \ldots = \partial^{r-2}u(0,x)/\partial t^{r-2} = 0$, $\partial^{r-1}u(0,x)/\partial t^{r-1} = \delta$ is called a *fundamental solution*. For homogeneous equations a fundamental solution corresponding to the problem above can be computed explicitly if $n = 1$, and application of the Radon method gives an analogous formula in the n-dimensional case.

Further developments of the methods discussed here are to be found in [2]–[6].

REFERENCES

[1]. Дикополов, Г. В., и Шилов, Г. Е., О корректных краевых задачах для уравнений в частных производных в полупространстве. *Изв. А. Н. С.С.С.Р., серия матем., 24*, 3 (1960), 369–380.

[2]. Паламодов, В. П., О корректных краевых задачах для уравнений в частных производных в полупространстве. *Изв. А. Н. С.С.С.Р., серия матем., 24*, 3 (1960), 381–386.

[3]. Дикополов, Г. В. и Шилов, Г. Е., О корректных краевых задачах в полупространстве для уравнений в частных производных с правой частью. *Сиб. матем. Жур., I*, 1, (1960), 45–61.

[4]. Шилов, Г. Е., О граничных задачах в четверти плоскости для уравнений в частных производных с постоянными коэффициентами. *Сиб. матем. Жур., II*, 1, (1961), 144–160.

[5]. Дикополов, Г. В., О краевых задачах для дифференциальных уравнений в частных производных с постоянными коэффициентами. *Успехи Матем. Наук, 16*, 1, (1961), 195–197.

[6]. Шилов, Г. Е., Уравнения с частными производными, Специальный курс, печатается в издательств Физматгиз (Москва).

SECOND ORDER ELLIPTIC EQUATIONS AND BOUNDARY VALUE PROBLEMS

By GUIDO STAMPACCHIA(1)

1. Introduction

While the theory of higher order elliptic differential equations with smooth coefficients grew up rapidly in the ten years since Gårding's paper [5] on the Dirichlet problem appeared, the development of the theory of equations with discontinuous coefficients has been very slow even for single second order equations.

The fact that solely a differential operator with smooth coefficients can be considered locally as a small perturbation of an operator with constant coefficients makes a significant difference. Consequently the techniques of Fourier transform and of singular integral operator, although extremely useful for equations with constant coefficients, fail in the case of equations with discontinuous coefficients.

The main reason for studying equations with discontinuous coefficients arises from nonlinear equations; however, there are several linear boundary value problems which escape the general theory of differential equations with smooth coefficients (for instance, the so-called transmission problem [26]).

The theory of equations with discontinuous coefficients is quite different in the case of two variables than it is for more variables. For two variables the theory commenced with a paper by C. B. Morrey [15] appearing in 1938 and developed with Nirenberg's results [21] in 1954 (see also [1]). For the applications of these results to nonlinear equations see the book of Miranda [14]. The theory of equations in two variables is closely connected with the beautiful and well developed theory of quasi-conformal mappings. See for expositions Courant-Hilbert [3] (the supplement to Chapter IV by L. Bers) and I. Vekua [29].

In more than two variables the theory was completely wrapped in mystery until a few years ago when the De Giorgi [4]–Nash [20] theorem was proved. This theorem is related to the equation in divergence form

$$(a_{ij}u_{x_i})_{x_j}=0 \tag{1.1}$$

with measurable and bounded coefficients when the condition

$$\nu^{-1}|\xi|^2\leqslant a_{ij}\xi_i\xi_j\leqslant\nu|\xi|^2 \quad (\nu\geqslant 1) \tag{1.2}$$

is satisfied; it states that any weak solution in $H^1(\Omega)$ is locally Hölder continuous.

This theorem led to many new results about the theory of linear and nonlinear second order elliptic equations in divergence form.

(1) This talk was prepared while the author was a Temporary Member at the Courant Institute of Mathematical Sciences, sponsored by the National Science Foundation, Contract No. NSF–GP98. I should like to thank the Courant Institute of Mathematical Sciences for making my trip to Stockholm possible.

A theory of more general equations in the form

$$a_{ij}u_{x_i x_j} + b_i u_{x_i} = 0 \tag{1.3}$$

with discontinuous coefficients is still lacking.

I want to limit myself to a review of some results for linear equations in divergence form and to mention only some applications to very special nonlinear problems. More general results are described in the talk by Nirenberg on nonlinear problems. Other questions on elliptic equations are described in an expository paper by Gilbarg [7].

2. Some notation

If Ω is a domain of the Euclidean space E^n, $\partial\Omega$ denotes its boundary and $\bar{\Omega}$ its closure. If $\xi \equiv (\xi_1, \xi_2, \ldots, \xi_n)$ is a vector in E^n we put $|\xi|^2 = \xi_1^2 + \ldots + \xi_n^2$. We shall say that a function $u(x)$ belongs to $C^m(\bar{\Omega})$ if u is continuous together with all the partial derivatives of order $\leqslant m$ in $\bar{\Omega}$ ($C^\infty(\Omega) = \bigcap_{m=0}^{\infty} C^m(\Omega)$). We shall denote by $C_\lambda^0(\bar{\Omega})$ the class of the functions which are Hölder continuous with exponent λ $(0 < \lambda \leqslant 1)$ in $\bar{\Omega}$, i.e.

$$\sup_{\substack{x', x'' \in \Omega \\ x' \neq x''}} \frac{|u(x') - u(x'')|}{|x' - x''|^\lambda} < +\infty .$$

A function of $C_1^0(\bar{\Omega})$ is also called a Lipschitz function in Ω.

Domains and boundary values of class C_λ^m are defined as usual.

The completion of $C^m(\bar{\Omega})$ with respect to the norm

$$\sum \|D^j u\|_{L^p(\Omega)} \quad (p \geqslant 1),$$

where $D^j u$ denotes any of the jth derivatives and the sum is extended to all derivatives of order $\leqslant m$ will be denoted by $H^{m,p}(\Omega)$, or more simply by $H^m(\Omega)$ when $p = 2$. By $H_0^{m,p}(\Omega)$ $[H_0^m(\Omega)]$ we shall denote the closure in $H^{m,p}(\Omega)$ $[H^m(\Omega)]$ of the set of the functions of $C^m(\Omega)$ vanishing near $\partial\Omega$. $u(x) \in H_{\mathrm{loc}}^{m,p}(\Omega)$ $[H_{\mathrm{loc}}^m(\Omega)]$ if $u(x) \in H^{m,p}(\Omega')$ for any compact subdomain Ω' of Ω. Consider the differential operator

$$M(u) = (a_{ij} u_{x_i})_{x_j}, \tag{2.1}$$

where a_{ij} are measurable and bounded functions defined in Ω and suppose it is uniformly elliptic, i.e.

$$\nu |\xi|^2 \leqslant a_{ij} \xi_i \xi_j \quad (\nu > 0). \tag{2.2}$$

A function $u(x)$ of $H^1(\Omega)$ is an M-subsolution [M-supersolution] in Ω, if

$$\int_\Omega a_{ij} u_{x_j} \phi_{x_j} dx \leqslant 0 [\geqslant 0] \tag{2.3}$$

for all ϕ of $H_0^1(\Omega)$ such that $\phi \geqslant 0$ a.e. in Ω.

A function $u(x)$ which belongs to $H_{\mathrm{loc}}^1(\Omega)$ and which satisfies (2.3) for every non-negative C^∞ function ϕ with compact support in Ω, will be called a local M-subsolution [or M-supersolution].

Consequently one defines the solutions and the local solutions.

3. The maximum principle

One of the best known results for second order elliptic equations is the Hopf maximum principle [8], (see also [3], p. 326) which holds for equations in the general form (1.3) even if the coefficients are supposed only to be bounded, but for smooth solutions. Different forms of the maximum principle which hold for very weak solutions, but for equations with smooth coefficients, have been found by E. Calabi [2] and W. Littman [12]. A balanced form of the maximum principle holds for M-subsolutions or for local M-subsolutions when the operator M has the form (2.1) and satisfies (2.2).

More generally we shall consider a continuous function $u(x)$ which belongs to $H^1_{\mathrm{loc}}(\Omega)$ and satisfies the inequality

$$\int_\Omega a_{ij} u_{x_i} \phi_{x_j} dx \leqslant \int_\Omega f_j \phi_{x_j} dx, \tag{3.4}$$

$$\left[\int_\Omega a_{ij} u_{x_i} \phi_{x_j} dx \geqslant \int_\Omega f_j \phi_{xj} dx\right] \tag{3.4'}$$

for every non-negative C^∞ function with compact support in Ω, where a_{ij} are measurable and bounded functions in Ω satisfying (2.2) and where $f_j \in L^p(\Omega)$ with $p > n$.

(3.4) [(3.4′)] means that u satisfies locally the differential inequality

$$M(u) \geqslant f_{jx_j} \quad [\leqslant]. \tag{3.5) [(3.5')}$$

Then we have the following weak form (for a strong form see § 4) of the maximum principle:

THEOREM I. *If M is an elliptic operator satisfying* (2.2), *if* $u(x) \in C^0(\bar{\Omega}) \cap H_{\mathrm{loc}}(\Omega)$ *and satisfies locally* (3.5) [(3.5′)] *with f_j in $L^p(\Omega)$ $(p > n)$, then there exists a constant $C(p,n)$ depending only on p and n such that*

$$\max_\Omega u(x) \leqslant \max_{\partial\Omega} u(x) + \frac{C(p,n)}{\nu} \sum \|f_j\|_{L^p(\Omega)} (\mathrm{mes}\ \Omega)^{1/n - 1/p}, \tag{3.6}$$

$$\left[\min_\Omega u(x) \geqslant \min_{\partial\Omega} u(x) - \frac{C(p,n)}{\nu} \sum \|f_j\|_{L^p(\Omega)} (\mathrm{mes}\Omega)^{1/n-1/p}\right]. \tag{3.6'}$$

We cannot give here the proof of the theorem, but we want to mention that the proof is based on the same idea introduced by Marcinkiewicz and used extensively by Zygmund [31], Hörmander [9] and others which consists in looking for information about the behaviour of the function

$$\phi(k) = \mathrm{mes}\{x \mid u > k\}.$$

Here we can prove that for $h > k > \max_{\partial\Omega} u$ the function ϕ satisfies an inequality of the type

$$\phi(h) \leqslant \frac{\mathrm{const}}{(h-k)^\alpha} [\phi(k)]^\beta, \tag{3.7}$$

where α, β are positive constants and $\beta > 1$ for $p > n$.

By iteration, one obtains from (3.7) that $\phi(d)=0$ where d is the right term of (3.6).

Exactly the same proof shows that the (3.6) [(3.6′)] holds for M-subsolutions [M-supersolutions] in Ω also if they are not continuous. The meaning of $\max_{\partial\Omega}$ and $\min_{\partial\Omega}$ has to be taken in a weak sense. Let $u(x)$ be a function belonging to $H^1(\Omega)$; $u(x)$ is bounded from above on $\partial\Omega$ by a constant Φ if there exists a sequence of functions $u_m \in C^1(\bar{\Omega})$ such that $u_m \leqslant \phi$ on $\partial\Omega$ and u_m tends to u in $H^1(\Omega)$. The smallest of such numbers Φ is denoted by $\max_{\partial\Omega} u$.

The same method of proof, as described above, can be used also for more general linear equations in order to obtain *a priori* bounds for the L^q-norm of the solution in terms of the L^p-norm of f_j and the L^2-norm of the solutions themselves; here $1/q=1/p-1/n$ if $p<n$ and $q=+\infty$ if $p>n$ (see Stampacchia [23, 24, 27*a*] and Maz'ya [13]). Ladyzenskaya and Uralt'seva [10, 11] and myself [27] used the same method in order to prove the boundedness of the solutions of some regular integrals of the calculus of variations.

About the Theorem I, Weinberger [30] proved that the constant $C(p,n)$ in (3.6) and in (3.6′) is given by

$$C(p,n)=\omega_n^{-1/n}\left(\frac{p-1}{p-n}\right)^{1-1/p} n^{1/n-1/p},$$

where ω_n is the measure of the unit n-sphere. Equality occurs in (3.6) and (3.6′) for Laplace's equation when Ω is a sphere.

4. Harnack's inequality and the Hölder continuity in the interior

The extension of the classic theorem by Harnack on the positive harmonic functions to the solution of more general elliptic equations has been given for $n=2$ by Bers and Nirenberg [1] and in general by Serrin [22] supposing, when $n>2$, the coefficients continuous.

Recently Moser [18] was able to prove the Harnack inequality for local solutions of (1.1) when the condition (1.2) is satisfied. The quoted paper by Bers and Nirenberg is based on a famous theorem by Lebesgue on the continuity of monotone functions with finite Dirichlet integrals. Such a result does not hold for functions in more variables. The Moser proof is based on a very different argument, and makes use of a special case of a theorem by John and Nirenberg [9*a*] which partially takes the place of Lebesgue's theorem.

The Moser theorem is the following

THEOREM II. *If u is a positive solution in Ω of equation* (1.1) *and* (1.2) *is satisfied, and if Ω' is a compact subdomain of Ω, then:*

$$\max_{\Omega'} u \leqslant c \min_{\Omega'} u, \tag{4.1}$$

where c depends on Ω', Ω and ν.

As a consequence of the Harnack inequality we can deduce the strong form of the maximum principle.

Consider now a positive solution of the equation

$$(a_{ij}u_{x_j})_{x_j} = f_{jx_j}, \tag{4.2}$$

where (1.2) is satisfied and $f_j \in L^p(\Omega)$ with $p > n$.

Using Moser's theorem and the maximum principle, as stated in Theorem I, we can get the following consequence. Let x_0 be any point of a compact subdomain Ω' of Ω and let

$$M(r) = \max_{|x-x_0|\leqslant r} u, \quad \mu(r) = \min_{|x-x_0|\leqslant r} u \quad (0 < r < \tfrac{1}{2}d),$$

where d denotes the distance from Ω' to $\partial\Omega$. Then, there exists a constant c depending on Ω', Ω, ν, such that

$$M(r) \leqslant c(\mu(r) + \sum \|f_j\|_{L^p(\Omega)} r^{1-(n/p)}). \tag{4.3}$$

From (4.3) it is possible to deduce easily that for any solution of (4.2) the following inequality holds

$$\omega\left(\frac{r}{2}\right) \leqslant \theta\omega(r) + C\sum \|f_j\|_{L^p(\Omega)} r^{1-(n/p)}, \tag{4.4}$$

where $\omega(r) = M(r) - \mu(r)$ is the oscillation of u in $|x - x_0| \leqslant r$ and where $\theta < 1$.

From (4.4) follows that any solution of (4.2) is Hölder continuous in the interior of Ω. When $f_j \equiv 0$ this result proves the De Giorgi–Nash theorem.

A direct proof of this theorem which does not use the Harnack theorem has been given by Moser [19].

5. Boundary value problems

The extension at the boundary of the latter result has been obtained independently by C. B. Morrey [16], Ladyzenskaya and Uralt'seva [10] and the author [25]. It states (in a special case)

THEOREM III. *If $u(x) \in H_0^1(\Omega)$ and satisfies the equation* (4.2) *where* (1.2) *holds and $f_j \in L^p(\Omega)$ with $p > n$ and if Ω satisfies a suitable condition $\mathcal{R}$, then $u(x)$ is Hölder continuous in $\bar{\Omega}$.*

The condition $\mathcal{R}$ on $\partial\Omega$ is the following [25]:

Let $B(y,r)$ be the sphere with center in y and radius r, there exist two constants K and r_0 such that for all $v \in C^1(B(y,r))$ vanishing on $\complement\Omega \cap B(y,r)$

$$|v(x)| \leqslant K \int_{B(y,r)} \frac{|\operatorname{grad} v|}{|x-t|^{n-1}} dt$$

for $x \in B(y,r)$, $y \in \partial\Omega$, $r < r_0$.

As a special case, condition $\mathcal{R}$ is satisfied if there exist two constants α, r_0 such that, for $y \in \partial\Omega$

$$\operatorname{mes}\{\complement\Omega \cap B(y,r)\} \geqslant \alpha \operatorname{mes} B(y,r) \quad (r < r_0)$$

(see [16] and [10]).

Using Hilbert space approach and the regularity theorem just mentioned we deduce that there exists one solution $u(x) \in C^0_\lambda(\bar{\Omega}) \cap H^1(\Omega)$ of the Dirichlet problem for the equation (4.2) where (1.2) holds, $f_j \in L^p(\Omega)$ with $p > n$ and for the boundary values

$$u - v \in H^1_0(\Omega),$$

provided that $v \in C^0_\lambda(\bar{\Omega}) \cap H^{1,q}(\Omega)$ $(q > n)$ and Ω satisfies the condition $\mathcal{R}$. We remark that if Ω is smooth the condition on v can be $v \in H^{1,q}(\Omega)$ with $q > n$.

Making use of the maximum principle of § 3 and the Harnack inequality of § 4 we can extend the Perron method from Laplace equation to the more general equation (1.1) satisfying (1.2). Using the preceding theorem, the existence of a barrier at the boundary can be proved if the condition $\mathcal{R}$ on Ω is satisfied.

Then *the Dirichlet problem for equation* (1.1) *satisfying* (1.2) *with arbitrarily assigned continuous boundary values has one solution* $u(x)$ *in* $C^0(\bar{\Omega}) \cap H^1_{\mathrm{loc}}(\Omega)$ *if the condition* $\mathcal{R}$ *on* Ω *is satisfied.*

We shall say that the points of $\partial\Omega$ are regular points for the operator M if the Dirichlet problem for arbitrarily assigned continuous boundary values is possible in $C^0(\bar{\Omega}) \cap H^1_{\mathrm{loc}}(\Omega)$.

The condition $\mathcal{R}$ assures that the points of $\partial\Omega$ are regular for the operator M.

We do not know if the class of domains which are regular for Laplace's operator coincides with the one of domains which are regular for the operator M given by (1.1) as is the case for operators with smooth coefficients [7a, 21 a, 28](1).

6. Some special cases of non-linear problems

The De Giorgi–Nash theorem and its extensions led to very interesting results on the differentiability of weak solutions of nonlinear equations or of minimizing functions of multiple integrals of the calculus of variations.

The first results in this direction were found by De Giorgi [4] for problems of the form

$$I(u) = \int_\Omega f(p)dx = \min \quad (p = \operatorname{grad} u),$$

supposing that

$$\nu^{-1}|\xi|^2 \leqslant f_{p_i p_j}(p)\xi_i\xi_j \leqslant \nu|\xi|^2 \quad (\nu \geqslant 1).$$

More general results have been obtained by C. B. Morrey [17] and Ladyzenskaya and Uralt'seva [10, 11] for general multiple integrals

(1) *Added in proof.* Recently H. F. Weinberger, W. Littman, and the author proved (Regular points for elliptic equations with discontinuous coefficients, to appear in *Ann. Scuola Norm. Sup., Pisa*) that a domain is regular for the operator M if and only if it is regular for the Laplace operator.

$$\int_\Omega F(x, u, p)\,dx.$$

These results require that F satisfy a condition of the type

$$\nu^{-1}(1+|p|^2)^\tau|\xi|^2 \leqslant F_{p_ip_j}(x, u, p)\xi_i\xi_j \leqslant \nu(1+|p|^2)^\tau|\xi|^2$$

with $\tau > -\frac{1}{2}$.

To begin with we remark that the classical variational problem of the minimal surface, where $F=\sqrt{1+|p|^2}$ is not included in these results. Recently it has been proved [27] that there exists a Lipschitz function in $\bar{\Omega}$ solution of the variational problem for the integral

$$I(u) = \int_\Omega f(p)\,dx = \min, \tag{6.1}$$

provided that f is strictly convex, i.e.

$$f_{p_ip_j}(p)\xi_i\xi_j > 0 \quad \text{for } \xi \neq 0,$$

and Ω is strictly convex and the boundary values are sufficiently smooth. The solution is analytic if the data are analytic. A similar theorem, using a different approach has been proved, independently by Gilbarg [6].

Such a statement may fail when F depends on the function u too. For the special integrals

$$I(u) = \int_\Omega \{f(p) + G(x, u)\}\,dx, \tag{6.2}$$

supposing that

$$f_{p_ip_j}(p)\xi_i\xi_j \geqslant \nu(1+|p|^2)^\tau|\xi|^2 \quad (\nu > 0, -\tfrac{1}{2} < \tau \leqslant 0)$$

and Ω strictly convex has been proved [27] the existence of a Lipschitz function minimizing $I(u)$ on the class of all Lipschitz functions vanishing on $\partial\Omega$ provided that suitable conditions on $G(x,u)$ are satisfied. Also here the solutions are smooth if the data are sufficiently smooth.

The assumption on the function $G(x,u)$ can be, for instance, the following:

$$\varliminf_{|u|\to+\infty} \frac{uG_u(x,u)}{|u|^\alpha} > -\nu\Lambda(\alpha, \Omega), \tag{6.3}$$

where $\alpha = 2(\tau+1)$ and $\Lambda\ (\alpha, \Omega)$ is given by

$$\Lambda(\alpha, \Omega) = \inf_{u\in H_0^{1,\alpha}(\Omega)} \frac{\int_\Omega |\operatorname{grad} u|^\alpha dx}{\int_\Omega |u|^\alpha dx}.$$

In particular $\Lambda(2,\Omega)$ is the first eigenvalue of the boundary value problem: $\Delta u + \lambda u = 0$ in Ω, $u=0$ on $\partial\Omega$.

Application of these theorems to the boundary value problems for the Euler equations of (6.1) and (6.2) leads to new results.

REFERENCES

[1]. BERS, L. & NIRENBERG, L., On linear and nonlinear boundary value problems in the plane. *Atti Convegno Internazionale sulle Equazioni a Derivate Parziali* (Trieste, 1954), 141–167.

[2]. CALABI, E., An extension of E. Hopf's maximum principle with an application to Riemannian geometry. *Duke Math. J.*, 25 (1958), 45–56.

[3]. COURANT, R., & HILBERT, D., *Methods of Mathematical Physics*, vol. II. Interscience Publishers, New York, 1962.

[4]. DE GIORGI, E., Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari. *Mem. Accad. Sci. Torino Cl. Sci. Fis. Mat. Nat.* (3), 3 (1957), 25–43.

[5]. GÅRDING, L., Dirichlet's problem for linear elliptic partial differential equations. *Math. Scand.*, 1 (1953), 55–72.

[6]. GILBARG, D., Boundary value problems for nonlinear elliptic equations in n variables. To appear in *Symposium on Nonlinear Problems*. Madison, Wisconsin, April 30–May 2, 1962.

[7] —— Some local properties of elliptic equations. *Proc. Symp. Pure Math.*, vol. 4, 127–140. Amer. Math. Soc., Providence, R.I., 1961.

[7*a*]. HERVÉ, R. M., Recherches axiomatiques sur la théorie des fonctions surharmoniques et du potentiel. *Ann. Inst. Fourier*, 12 (1962), 415–571.

[8]. HOPF, E., Elementare Bemerkungen über die Lösungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus. *Sitzungsberichte, Berlin Akad. Wiss.*, 19 (1927), 147–152.

[9]. HÖRMANDER, L., Estimates for translation invariant operators in L_p spaces. *Acta Math.*, 104 (1960), 93–140.

[9*a*]. JOHN, F. & NIRENBERG, L., On functions of bounded mean oscillation. *Comm. Pure Appl. Math.*, 14 (1961), 415–426.

[10]. LADYZENSKAYA, O. A. & URALT'SEVA, N. N., Quasi-linear elliptic equations and variational problems with many independent variables. *Russian Math. Surveys, London Math. Soc.*, (16) 1 (1961), 17–91.

[11]. —— On the smoothness of weak solutions of quasi-linear equations in several variables and of variational problems. *Comm. Pure Appl. Math.*, 14 (1961), 481–495.

[12]. LITTMAN, W., A strong maximum principle for weakly L-subharmonic functions. *J. Math. Mech.*, 8 (1959), 761–770.

[13]. MAZ'YA, V. G., Some estimates for solutions of elliptic second-order equations. *Dokl. Akad. Nauk S.S.S.R.*, 137 (1961); translation in English, 413–416.

[14]. MIRANDA, C., Equazioni alle derivate parziali di tipo ellittico. *Ergebnisse der Mathematik und ihrer Grenzgebiete* (*N.F.*), Heft 2. Springer, Berlin, 1955.

[15]. MORREY, C. B., On the solutions of quasi-linear elliptic partial differential equations. *Trans. Amer. Math. Soc.*, 43 (1938), 126–166.

[16]. MORREY, C. B., JR., Second order elliptic equations in several variables and Hölder continuity. *Math. Z.* 72 (1959), 146–164.

[17]. —— Existence and differentiability theorems for variational problems for multiple integrals. *Conf. on Partial Differential Equations and Continuum Mechanics*. Univ. of Wisconsin Press, Madison, Wis., 1961.

[18]. MOSER, J., On Harnack's theorem for elliptic differential equations. *Comm. Pure Appl. Math.*, XIV (1961), 577–591.

[19]. —— A new proof of De Giorgi's theorem concerning the regularity problem for elliptic differential euqations. *Comm. Pure Appl. Math.*, 13 (1960), 457–468.

[20]. NASH, J., Continuity of the solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80 (1958), 931–954.

[21]. NIRENBERG, L., On non-linear elliptic partial differential equations and Hölder continuity. *Comm. Pure Appl. Math.*, 6 (1953), 103–156.

[21*a*]. OLĚINIK, O. A., On the Dirichlet problem for equations of elliptic type (Russ.) *Math. Sb. N. S.* 24 (66) (1949), 3–14.

[22]. SERRIN, J., On the Harnack inequality for linear elliptic equations. *J. Analyse Math.*, 4 (1955–1956), 292–308.

[23]. STAMPACCHIA, G., Contributi alla regolarizzazione delle soluzioni dei problemi al contorno per equazioni del secondo ordine ellittiche. *Ann. Scuola Norm. Sup. Pisa*, (III) 12 (1958), 223–244.

[24]. —— Régularisation des solutions de problèmes aux limites elliptiques à données discontinues. *Intern. Symp. Lin. Spaces*. Jerusalem, 1960.

[25]. —— Problemi al contorno ellittici, con dati discontinui, dotati di soluzioni hölderiane. *Ann. Mat. Pura Appl.*, (IV) 51 (1960), 1–38.

[26]. —— I problemi di trasmissione per le equazioni di tipo ellittico. *Semin. Matem. Bari*, N. 58, (1960).

[27]. —— On some regular multiple integral problems in the calculus of variations. (To appear; New York Report, 1962).

[27*a*]. —— Equations elliptiques à données discontinues. *Séminaire Schwartz* (1960–61).

[28]. TAUTZ, G., Zur Theorie der elliptischen Differentialgleichungen. *Math. Ann.*, 118 (1943), 733–770.

[29]. VEKUA, I., Generalized analytic functions. *Gos. Izd. Fiz.-Mat. Lit.*, Moscow, 1959.

[30]. WEINBERGER, H., Symmetrization in uniformly elliptic problems. (To appear in Studies in Math. An. and related topics: *Essays in honor of G. Polya.*)

[31]. ZYGMUND, A., On a theorem of Marcinkiewicz concerning interpolation of operations. *J. Math. Pures Appl.* 35 (1956), 223–248.

CONJUGATE HARMONIC FUNCTIONS IN SEVERAL VARIABLES

By E. M. STEIN

1. Introduction and background

The purpose of this note is to report on some recent progress in the theory of conjugate harmonic functions in several variables.

We begin by sketching the requisite background.

In the Euclidian n-dimensional space, E_n, we consider an n-tuple of functions $u_1, u_2, \ldots, u_n$ which satisfies the equations

$$\left.\begin{aligned} &\sum_{k=1}^{n} \frac{\partial u_k}{\partial x_k} = 0, \\ \text{and} \qquad &\frac{\partial u_k}{\partial x_j} = \frac{\partial u_j}{\partial x_k}. \end{aligned}\right\} \tag{1}$$

Clearly (1) is locally equivalent with the statement that the vector $u = (u_1, \ldots, u_n)$ is the gradient of a harmonic function H, that is $u = \nabla H$. That such a system might be a fruitful n-dimensional analogue of the usual notion of an analytic function (the case $n=2$) was suggested by various writers. This seems to be borne out particularly when one considers (1) in connection with problems of $n-1$ dimensional Fourier analysis.

For this purpose we distinguish one of the variables, say x_1, and consider the half space E_n^+ of points where $x_1 > 0$ and $(x_2, \ldots, x_n)$ are arbitrary. The subspace E_{n-1} of points $(x_2, \ldots, x_n)$ may be identified with the boundary hyper-plane $x_1 = 0$ of E_n^+.

Now let $f(x_2, \ldots, x_n)$ be an arbitrary (say L^2) function on E_{n-1}, and let $u(x_1, \ldots, x_n)$ be its Poisson integral. That is, u is harmonic in E_n^+ and takes on the boundary value f when $x_1 = 0$. Starting from u it is easy to pass to an n-tuple of functions in E_n^+ which satisfy (1), with $u_1 = u$, and so that $u_2, \ldots, u_n$ are also Poisson integrals of their (L^2) boundary values. Finally set $f_k(x_2, \ldots, x_n) = u_k(0, x_2, \ldots, x_n)$, $k = 2, \ldots, n$.

In this way we see that starting from an "arbitrary" function on E_{n-1} we can pass to a system of functions satisfying (1); this construction also leads to $n-1$ linear transformations of conjugacy

$$R_k : f \to f_k \quad (k = 2, \ldots, n). \tag{2}$$

The Riesz transforms, R_k, are the natural generalizations of the Hilbert transform. Their significance can most easily be understood in terms of the Fourier transform. Thus in these terms R_k is a multiplier transformation corresponding to the factor $ix_k(x_2^2 + x_3^2 \ldots x_n^2)^{-\frac{1}{2}}$; (see Horvath [3]).

A further significance of the R_k is that they are the basic building blocks for the "singular integral" transformations which have found wide application in partial differential equations; (see Calderon and Zygmund [1]). We add here an important inequality satisfied by the Riesz transforms:

$$\|R_k f\|_p \leqslant A_p \|f\|_p, \; 1 < p < \infty.$$

Another significant consequence of the notion (1) of conjugacy is the fact that $|u|^p=(u_1^2+\ldots u_n^2)^{p/2}$ is sub-harmonic for $p\geqslant(n-2)/(n-1)$. This, together with harmonic majorization, allows us to develop a theory of H^p spaces extending the classical notion of the Hardy spaces (when $n=2$). The basic result when $p=1$ is the following extension of a well-known theorem of F. and M. Riesz. If $d\mu$ is a finite measure on E_{n-1} and its $n-1$ Riesz transforms are also finite measures, then all n measures are absolutely continuous. For these results see Stein and Weiss [6].

Finally, it is possible to extend results of Plessner and Privalov dealing with boundary behavior of analytic functions to n-tuples satisfying (1). We may state the result as follows: Let E be a subset of E_{n-1} of positive measure. If for every $(\bar{x}_2,\ldots,\bar{x}_n)\in E$, lim u_1 exists as the point $(x_1,\ldots,x_n)$ converges to $(0,\bar{x}_2,\ldots,\bar{x}_n)$ non-tangentially, then the same is true for $u_2,\ldots,u_n$ for almost all points in E; the converse statement also holds; (see Stein [5]).

We shall now describe two new directions of investigation in this subject. The first reveals intimate connection with various classical expansions. The second is closely related to the theory of representations of the rotation group in n-variables.

2. Classical expansions

Our first point of departure is the consideration of harmonic functions which are radial in $n-1$ of the variables. This, as is well known, leads to problems analagous to Fourier analysis but for Bessel functions of order $(n-3)/2$, and by analogy to Bessel functions of real order $\geqslant -\frac{1}{2}$.(1) A similar situation holds for ultraspherical (Gegenbauer) expansions, and other classical expansions. Let us consider the matter in more detail. We set $x=x_1$ and $y=(x_2^2+\ldots x_n^2)^{\frac{1}{2}}$. Consider the system (1) in the half-space E_n^+, and let $u_1(x_1,\ldots,x_n)=u(x,y)$ be radial in the variables $x_2,\ldots,x_n$; suppose that it is the Poisson integral of an $f(x_2,\ldots,x_n)=f(y)$ defined on E_{n-1}. Then $u(x,y)=\partial H/\partial x$ where $H(x,y)$ is harmonic as a function of $x_1,\ldots,x_n$. Thus H satisfies the equation

$$\frac{\partial^2 H}{\partial x^2}+\frac{\partial^2 H}{\partial y^2}+\frac{n-2}{y}\frac{\partial H}{\partial y}=0. \tag{3}$$

As one might guess, there are now no longer $n-1$ distinct conjugates to $u(x,y)$, but only one. In fact

$$u_k(x,y)=\frac{x_k}{y}\frac{\partial H}{\partial y}\quad (k=2,\ldots,n).$$

Thus set $v(x,y)=\partial H/\partial y=(u_2^2+\ldots,u_n^2)^{\frac{1}{2}}$. Then u and v together satisfy the singular "Cauchy-Riemann equations"

$$\begin{aligned} u_y-v_x&=0,\\ u_x+v_y+\frac{n-2}{y}v&=0. \end{aligned} \tag{4}$$

(1) This is also the point of view of axially symmetric potential theory, see e.g. Weinstein [7].

Now replace the constant $n-2$ appearing in (3) and (4) by the positive constant 2λ. Then it is natural to expand $u(x,y)$ (as a function of y) as a Hankel integral involving the Bessel function of order $\lambda-\frac{1}{2}$; and its conjugate $v(x,y)$ in terms of the Bessel function of order $\lambda+\frac{1}{2}$. One might expect a variety of results for this type of conjugacy similar to those discussed in the preceding section; this turns out to be the case. We shall not go into further details, however, concerning the problem of expansions in Bessel functions; instead we shall discuss at some length the analagous problem involving ultraspherical expansions. In fact, the results that can be obtained for Bessel expansions are very similar to the results we shall now state for the ultraspherical case.

The problem for ultraspherical expansions arises, in the first place, by considering functions which are harmonic in the interior of the unit sphere in n-dimensions in terms of their boundary values on the surface of the sphere; and then, by restricting consideration to those boundary functions which are "radial" ("spherical" in the modern terminology) that is, depend only on the distance from the "north pole". The notion of conjugacy, (4), turns out to be the one which links the normal to the tangential derivatives of such harmonic functions. More particularly this leads to ultraspherical expansions of type $(n-2)/2$, and to conjugate expansions of type $n/2$. By analogy we pass to expansions of type λ, and to their conjugate expansions of type $\lambda+1$, $\lambda\geqslant 0$.

The ultraspherical polynomials of type λ, P_n^λ, are defined by

$$\sum_{n=0}^{\infty} r^n P_n^\lambda(\cos\theta) = (1-2r\cos\theta+r^2)^{-\lambda} \quad (\lambda>0).$$

Then P_n^λ are orthogonal with respect to the measure $(\sin\theta)^{2\lambda}d\theta$ on $(0,\pi)$. It is with respect to this measure that the L^p spaces are then defined.

Suppose now that $f(\theta)$ is an arbitrary function on $(0,\pi)$, and let it have the expansion $f(\theta)\sim\sum a_n P_n^\lambda(\cos\theta)$. We then define its "Poisson integral" $f(r,\theta)$ by

$$f(r,\theta)=\sum_{n=0}^{\infty} a_n r^n P_n^\lambda(\cos\theta) \quad (0\leqslant r<1), \tag{5}$$

and set $f(r,\theta)=u(x,y)$ in cartesian coordinates. Then u satisfies the singular Laplace equation (3) (with $n-2=2\lambda$), in the upper semidisc $x^2+y^2<1$, $y>0$. If we look for its conjugate function $v(x,y)$ given by (4) (again $n-2=2\lambda$), then we obtain

$$v(x,y)=\tilde{f}(r,\theta)=-2\lambda\sum_{n=1}^{\infty}\frac{a_n}{n+2\lambda}r^n\sin\theta P_{n-1}^{\lambda+1}(\cos\theta). \tag{6}$$

Finally, this leads us to consider the natural conjugacy mapping $f(\theta)\to\tilde{f}(\theta)$, where

$$\tilde{f}(\theta)\sim -2\lambda\sum_{n=1}^{\infty}\frac{a_n}{n+2\lambda}\sin\theta P_{n-1}^{\lambda+1}(\cos\theta).$$

Some of the results obtained may be listed as follows:

(*a*) The inequality $\|\tilde{f}\|_p\leqslant A_p\|f\|_p$, $1<p<\infty$, which is a generalization of the classical conjugacy inequality of M. Riesz.

(*b*) A maximum principle for those functions u which are "subharmonic" in the sense that

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{2\lambda}{y}\frac{\partial u}{\partial y}\geqslant 0.$$

The regions considered may contain the singular line $y=0$. Because of this singularity the usual Hopf maximum principle is not applicable; but the situation can be remedied by limiting oneself to functions even in the y variable.

(*c*) If u and v satisfy (4) (with $n-2=2\lambda$) then $(u^2+v^2)^{p/2}$ is "subharmonic" if $p\geqslant(2\lambda)/(2\lambda+1)$.

(*d*) From (*b*) and (*c*) one can obtain a theory of H^p spaces, for ultraspherical expansions, whenever $p\geqslant(2\lambda)/(2\lambda+1)$. In fact if $F(re^{i\theta})=f(r,\theta)+i\tilde{f}(r,\theta)$, (see equations 5 and 6), and $\sup_{r<1}\int_0^\pi|F(re^{i\theta})|^p\sin^{2\lambda}\theta d\theta<\infty$, $p>(2\lambda)/(2\lambda+1)$, then $\lim_{r\to 1}F(re^{i\theta})$ exists almost everywhere, in the L^p norm, and dominatedly. Incidentally, (*a*) may be viewed as describing the H^p spaces, when $p>1$.

(*e*) The results of the H^p theory, when $p=1$, may be restated as follows. Suppose $d\mu_1$ and $d\mu_2$ are two measures on $(0,\pi)$ which are finite, in the sense that $\int_0^\pi\sin^{2\lambda}\theta|d\mu_i(\theta)|<\infty$. Let

$$d\mu_1\sim\sum a_n P_n^\lambda(\cos\theta)$$

and

$$d\mu_2\sim 2\lambda\sum\frac{a_n}{n+2\lambda}\sin\theta P_{n-1}^{\lambda+1}(\cos\theta).$$

Then both $d\mu_1$ and $d\mu_2$ are absolutely continuous.

The results sketched in this section, and others of this type, were obtained jointly with B. Muckenhoupt.

3. Representations of the rotation group

The second general area we shall deal with concerns the study of other systems, akin to (1), which are in a natural sense generalizations of the Cauchy-Riemann equations.

The results are closely connected to the theory of representations of the rotation group in n variables. Here we shall limit ourselves to the case of three variables, where the problems raised can be solved explicitly. While the situation for any number of variables is more complex, a similar outcome for the general case is indicated.

We begin by giving examples of some of the systems, in addition to (1), which we shall study in a general setting. Let us use the notation

$$\partial=\frac{\partial}{\partial x_1}+i\frac{\partial}{\partial x_2},\quad \bar{\partial}=\frac{\partial}{\partial x_1}-i\frac{\partial}{\partial x_1},\quad \partial_3=\frac{\partial}{\partial x_3}.$$

A system considered by Moisil and Theodoresco (see [4]) consists of two *complex*-valued functions u and v satisfying the equations

$$\begin{aligned}\bar{\partial}u+\partial_3 v&=0,\\ \partial v-\partial_3 u&=0.\end{aligned}\tag{7}$$

It may be of interest to point out that (7) is merely the electron equation of Dirac in the case without external force, with zero mass, and independent of time.

It should also be observed that passing to the real and imaginary components of u and v, (7) gives four equations in these four components. If we had taken one of these components to be zero (e.g. the imaginary part of u), then we recover essentially the gradient system (1), when $n=3$. Thus it is natural to raise the same problems about system (7) as we raised in the case of system (1). In particular we may ask; for what $p>0$, is $(|u|^2+|v|^2)^{p/2}$ sub-harmonic, when u and v satisfy (7). The answer is $p\geqslant\frac{1}{2}$; this is identical with the result for system (1), when $n=3$. Thus, by what was said above this conclusion actually extends our previous result for (1), when $n=3$.

Another variant of (1) occurs when we take a given harmonic function H and form its gradient of order l, l a positive integer. Thus $u=\nabla^l H$, and u is actually a tensor of rank l whose components consist of the various lth. derivatives of H. Such tensors u can also be characterized by a system of equations like (1). If $|u|$ denotes the usual Euclidean norm of this tensor, we may again ask when $|u|^p$ is sub-harmonic. It was shown by Calderon and Zygmund (1) that $|u|^p$ is sub-harmonic when $p\geqslant 1/(l+1)$. We should note that the larger the system becomes in terms of the number of components and relations between them, the stronger the result for sub-harmonicity that obtains.

From the examples cited above the following two general problems arise: (*a*) to characterize all possible natural generalizations of the Cauchy-Riemann equations; (*b*) to find the best exponent p so that the analogue of $|u|^p$ is sub-harmonic, in each given case.

To study problem (*a*) we have to consider the invariance properties of the examples cited. These systems are trivially invariant under the group of translations. They are distinguished, however, by their transformation properties under rotations. In fact, if the underlying space is transformed according to the group of rotations, each system is naturally transformed according to an appropriate representation of this group. (2) In the case of (1) this representation is the principal one—which assigns to each rotation itself. This is most easily seen by writing $u=\nabla H$. In the case of the gradient of order l, $u=\nabla^l H$, the representation is the one induced on spherical harmonics of degree l. Alternatively this representation, which is irreducible, can be described as acting under symmetric tensors of rank l, all of whose traces vanish. Finally, a study of example (7) shows that we must include the two-valued as well as the single-valued representations of the rotation group. In fact let R_3 denote the 3-dimensional rotation group, and U its "spinor" group; i.e. U is the two-dimensional group of unitary matrices of determinant one. The group R_3 is not simply-connected, but has U as its simply connected covering group. Thus U has only single-valued representation, and these are in one-one correspondence with the totality of single-valued or double-valued representations of R_3. The system (7) then corresponds to the principal representation of U, which is the usual spinor representation of R_3.

To state our general results we shall need some further notation. Let r

(1) This result is unpublished.

(2) For a survey of the theory of representations of the 3-dimensional rotation group see Gel'fand and Shapiro [2].

denote a positive integer, and let $l=r/2$. Thus $l=0,\frac{1}{2},1,\ldots$ For each such l, let V_l denote the (complex unitary) vector space of symmetric tensors of rank r whose indices take on the values 1 or 2. Since the group U consists of 2×2 matrices which are unitary, there is a natural representation of U on V_l; and thus there is a natural (possibly two-valued) irreducible representation of R_3 on V_l. Let $\varrho\to R^l(\varrho)$ denote this representation, $\varrho\in R_3$. In fact, $R^l(\varrho)$ is single-valued if l is integral, and double-valued otherwise. This scheme describes all the irreducible representations of R_3.

We return to our first problem. To each l we shall associate a system, S_l, generalizing (1). Let $u(x)$ be a function on E_3 whose values are in V_l; the differential equations determining the system are, in components,

$$(S_l)\qquad \begin{cases} \bar{\partial} u_{1i_2i_3\ldots i_r}+\partial_3 u_{2i_2i_3\ldots i_r}=0, \\ \partial u_{2i_2i_3\ldots i_r}-\partial_3 u_{1i_2i_3\ldots i_r}=0. \end{cases}$$

It may be shown that the system S_l transforms according to the representation R^l in the following sense. If u satisfies (S_l), then $v(x)=R^l(\varrho)[u(\varrho^{-1}(x))]$ satisfies (S_l) also, for each $\varrho\in R_3$.

Therefore we see, in answer to the first problem raised, how to assign in a natural way a system to each irreducible representation of R_3. It is desirable to describe the system S_l in a more intrinsic fashion, which will also indicate its n-dimensional generalization.

We had $u(x)\in V_l$; therefore its gradient $\nabla u(x)$ belongs to $V_l\otimes V_1$. ∇u transforms according to the representation $R^l\otimes R^1$, ($\otimes$ denotes the tensor product). However the tensor product representation is not irreducible; it reduces into a direct sum of irreducible ones. Let $V_{l,1}$ denote the irreducible sub-space of $V_l\otimes V_1$ containing the *highest* weight vector. Then $R^l\otimes R^1$ restricted to $V_{l,1}$ is equivalent with R^{l+1}. Incidentally, this gives us the so-called Cartan composition of R^l with R^1.

The content of equations (S_l) is exactly the statement that $\nabla u(x)$ belongs to the sub-space $V_{l,1}$ of $V_l\otimes V_1$.

It is to be observed that when $l=\frac{1}{2}$, (S_l) reduces immediately to the system (7). When $l=1$, then after a simple change of basis, (S_l) becomes the system (1). Similarly for any integral l this system can be shown to be locally characterized as the gradient of order l of a single harmonic function.

We come now to the second question raised earlier. For what p is $|u|^p$ sub-harmonic(1) when u satisfies the system (S_l)?

The result is as follows: When l is integral, then $|u|^p$ is subharmonic if $p\geqslant 1/(l+1)$. When l is half-integral, then $|u|^p$ is sub-harmonic when $p\geqslant 1/(l+\frac{3}{2})$. Both these conclusions are best possible. The interest of this result is that it allows us to extend the theory of H^p spaces to the systems (S_l) for the above mentioned range of p's.

The results in this section were obtained in collaboration with G. Weiss.

(1) Recall that V_e is a unitary vector space, and hence the norm $|u|$ is defined on it.

REFERENCES

[1]. CALDERON, A. P. & ZYGMUND, A., Singular integral operators and differential equations. *Amer. J. Math.*, 79 (1957), 901–921.

[2]. GEL'FAND, I. M. & SHAPIRO, Z. YA., Representations of the group of rotations in 3-dimensional space and their applications. *Amer. Math. Soc. Transl., series* 2, 2 (1956), 207–316. Translated from Uspehi Mat. Nauk 47 (1952), 3–117.

[3]. HORVATH, J., Sur les fonctions conjuguées à plusieurs variables *Nederl. Akad. Wetensch. Proc. Ser. A.*, 16 (1953), 17–29.

[4]. MOISIL, G. C. & THEODORESCO, N., Fonctions holomorphes dans l'espace, *Mathematica (Cluj)* V (1931), 142–159.

[5]. STEIN, E. M., On the theory of harmonic functions of several variables. II. *Acta Math.*, 106 (1961), 137–174.

[6]. STEIN, E. M. & WEISS, G., On the theory of harmonic functions of several variables. I. *Acta Math.*, 103 (1960), 25–62.

[7]. WEINSTEIN, A., Generalized axially symmetric potential theory, *Bull. Amer. Math. Soc.*, 59 (1953), 20–38.

THE "OUTER FUNCTIONS" AND THEIR ROLE IN FUNCTIONAL CALCULUS

By BÉLA SZ.-NAGY

1. Let H_p denote the Hardy class of order p $(1 \leqslant p \leqslant \infty)$, of functions defined in the interior D of the unit circle C; in particular H_∞ consists of the functions which are holomorphic and bounded in D; we have

$$H_1 \supset H_p \supset H_\infty \quad (1 < p < \infty).$$

For $\varphi(z) \in H_1$, the limit $\varphi(e^{i\theta}) = \lim_{r \to 1} \varphi(re^{i\theta})$ exists a.e. on C, and

$$\int_0^{2\pi} \varphi(e^{i\theta})\, d\theta = \lim_{r \to 1} \int_0^{2\pi} \varphi(re^{i\theta})\, d\theta = 2\pi\varphi(0)$$

[Fatou, F. Riesz]. If $\varphi(z) \not\equiv 0$, then $\log|\varphi(e^{i\theta})| \in L(0, 2\pi)$ [Szegő].

Every function $\varphi(z) \in H_1$, $\varphi(z) \not\equiv 0$, can be factorised into the product of an "inner factor" $\varphi_0(z)$ and an "outer factor" $\varphi_1(z)$ by putting

$$\varphi_1(z) = \exp\left[\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log|\varphi(e^{i\theta})|\, d\theta\right];$$

we have $\qquad |\varphi_0(z)| \leqslant 1 \text{ in } D, \quad |\varphi_0(e^{i\theta})| = 1 \text{ a.e. on } C,$

and $\qquad \varphi_1(z) \in H_1 \quad$ [Szegő].

The function $\varphi(z)$ is called "inner" or "outer" if its outer or inner factor is constant, respectively. Thus $\varphi(z)$ is outer if

$$\varphi(z) = K \exp\left[\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log|\varphi(e^{i\theta})|\, d\theta\right], \; |K| = 1,$$

or, equivalently, $\quad -\infty < \log|\varphi(0)| = \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} \log|\varphi(e^{i\theta})|\, d\theta.$

Beurling [1] has found, in connection with some "closure" and "extinction" problems in Hilbert space, the following important result:

(I) *For $\varphi(z) \in H_2$, the set of functions $\{z^n \varphi(z)\}$ spans H_2 with respect to the norm*

$$\|\varphi\|_2 = \left\{\frac{1}{2\pi} \int_0^{2\pi} |\varphi(e^{i\theta})|^2 d\theta\right\}^{\frac{1}{2}}$$

if, and only if, φ is an outer function.

Since H_2 is a Hilbert space with respect to this norm, this result is equivalent with the following:

(I′) *A function $\varphi(z) \in H_2$ is an outer function if, and only if, there exists no function $\psi(z) \in H_2$ such that*

$$(z^n\varphi, \psi) \equiv \int_0^{2\pi} e^{in\theta}\varphi(e^{i\theta})\overline{\psi(e^{i\theta})}\,d\theta = 0 \quad (n=0, 1, \ldots),$$

except the function $\psi(z) \equiv 0$.

This result has the following counterpart for the class H_∞, found by Sz.-Nagy and Foiaş [3].

(II) *A function* $\varphi(z) \in H_\infty$ *is an outer function if, and only if, there exists no function* $p(\theta)$ *with the properties*

$$p(\theta) \geqslant 0, p(\theta) \in L(0, 2\pi), \quad \int_0^{2\pi} e^{in\theta}\varphi(e^{i\theta})p(\theta)\,d\theta = 0 \quad (n=0, 1, \ldots), \tag{1}$$

except the function $p(\theta) = 0$ *a.e.*

2. The outer functions have evidently no zeros in D, but this property does not characterise them. E.g. the function $\varphi(z) = \exp((z+1)/(z-1))$ belongs to H_∞, has no zeros in D, but it is not an outer function since $|\varphi(e^{i\theta})| = 1$ a.e., hence its outer factor is $\varphi_1(z) \equiv 1$.

A characterisation of the outer functions in the class H_∞ by their zeros is nevertheless possible if we first extend their definition in the following way.

We consider linear operators T in Hilbert space, such that $\|T\| \leqslant 1$, i.e. *contraction operators*. The underlying Hilbert space may be of arbitrary finite or infinite dimension, we suppose only that it be non trivial, i.e. that it consist not only of the zero vector. We suppose further that the contraction operator T be *absolutely continuous* in the sense that it admits a (strong) unitary dilation U_T whose spectral measure dE_θ is absolutely continuous with respect to Lebesgue measure.

Let $\tilde{D}$ denote the set of *all* such operators T. By a result of Sz.-Nagy–Foiaş [2], $\tilde{D}$ contains in particular all contraction operators which are *completely non unitary*, i.e. which are not unitary and have no unitary part.

For every function $\qquad \varphi(z) = \sum_{n=0}^{\infty} c_n z^n \in H_\infty$

we can define the operator $\varphi(T)$ for $T \in \tilde{D}$ by Abel summation of the operator series

$$\sum_{c=0}^{\infty} c_n T^n,$$

namely by putting

$$\varphi_r(T) = \sum_{n=0}^{\infty} c_n r^n T^n \quad (0 < r < 1) \quad \text{and} \quad \varphi(T) = \lim_{r\to 1} \varphi_r(T).$$

The series of $\varphi_r(T)$ converges in norm since

$$\| c_n r^n T^n \| \leqslant | c_n | r^n,$$

and the limit of $\varphi_r(T)$ for $r \to 1$ exists in the sense of strong operator convergence. In fact, we have

$$\varphi_r(T) = \mathrm{pr} \sum_{n=0}^{\infty} c_n r^n U_T^n = \mathrm{pr} \int_0^{2\pi} \sum_{n=0}^{\infty} c_n r^n e^{in\theta} dE_\theta = \mathrm{pr} \int_0^{2\pi} \varphi(re^{i\theta})\,dE_\theta$$

and
$$\int_0^{2\pi} \varphi(re^{i\theta})\,dE_\theta \to \int_0^{2\pi} \varphi(e^{i\theta})\,dE_\theta = \varphi(U_T) \quad \text{strongly}$$

by Lebesgue's theorem, since $\varphi(re^{i\theta})$ converges to $\varphi(e^{i\theta})$ boundedly and a.e. with respect to Lebesgue measure, and hence a.e. with respect to the spectral measure dE_θ too. Hence $\varphi(T)$ exists and we have

$$\varphi(T) = \operatorname{pr} \varphi(U_T).$$

In particular, if $|z| < 1$, the operator $T_z = z\,I$ (where I denotes the identity operator of the Hilbert space) is a completely non unitary contraction, hence belongs to the set $\tilde{D}$, and we have

$$\varphi(T_z) = \varphi(z) \cdot I.$$

Thus if we identify each complex number c with the operator T_c, we can consider the above definition of $\varphi(T)$ as a natural extension of the function φ from the domain D to the larger domain $\tilde{D}$.

This extension has many useful properties, such as additivity and multiplicativity with respect to φ for fixed T, and the inequality

$$\|\varphi(T)\| \leqslant \sup_D |\varphi(z)|,$$

thus we obtain a reasonable and useful functional calculus for T.

3. Moreover, this extension makes possible the following characterisation of the outer functions (Sz.-Nagy–Foiaş [3]):

(III) *A function $\varphi(z) \in H_\infty$ is outer if, and only if, its extension to the domain $\tilde{D}$ has no zeros, i.e. if $\varphi(T)$ is different from the zero operator for every absolutely continuous contraction operator T; even it suffices that $\varphi(T)$ be different from* 0 *for every completely non unitary contraction operator T.*

Proof. (*i*) Suppose $\varphi(z) \in H_\infty$ and $\varphi(T) = 0$ for some $T \in \tilde{D}$. Then we have for every element $f \neq 0$ of the Hilbert space (in which the operator T acts)

$$0 = (T^n \varphi(T) f, f) = (U_T^n \varphi(U_T) f, f) = \int_0^{2\pi} e^{in\theta} \varphi(e^{i\theta})\, d(E_\theta f, f)$$
$$= \int_0^{2\pi} e^{in\theta} \varphi(e^{i\theta}) p(\theta)\, d\theta \quad (n = 0, 1, \ldots)$$

with
$$p(\theta) = d(E_\theta f, f)/d\theta,$$

and
$$\int_0^{2\pi} p(\theta)\, d\theta = \int_0^{2\pi} d(E_\theta f, f) = (f, f) \neq 0.$$

Thus $p(\theta)$ has all the properties (1) without vanishing a.e., thus, by (II), $\varphi(z)$ can not be an outer function.

(ii) We suppose now that $\varphi(z)$ is not an outer function, and show that its extension $\varphi(T)$ has some zero T in $\tilde{D}$, even a completely non unitary contraction operator T. Obviously, we can suppose that $\varphi(z) \not\equiv 0$.

By (II) there exists a function $p(\theta)$ with the properties (1) and not vanishing a.e., i.e. with $\int_0^{2\pi} p(\theta) d\theta > 0$. Then

$$f_0(\theta)=\sqrt{p(\theta)}\in L_2\,(0,2\pi),\quad (f_0,f_0)>0.$$

We introduce the notation V for the operator of multiplication by $e^{i\theta}$ of the functions $f(\theta)$ in $L_2(0,2\pi)$; V is a unitary operator in $L_2(0,2\pi)$ with absolutely continuous spectral measure. Hence $\varphi(V)$ exists; in fact $\varphi(V)$ is equal to the multiplication by the function $\varphi(e^{i\theta})$.With these notations, we can express the above hypotheses as follows:

$$(V^n\varphi(V)f_0,f_0)=0\quad (n=0,1,\ldots);\quad (f_0,f_0)>0. \tag{2}$$

It follows that $V^n\varphi(V)f_0$ is orthogonal also to the elements $V^{k*}f_0$ $(k=0,1,\ldots)$ and hence also to the subspace spanned by them, i.e. to

$$\mathfrak{M}=\bigvee_{k=0}^{\infty} V^{*k}f_0.$$

Let P denote the operator of the orthogonal projection to this subspace $\mathfrak{M}$, and consider the subspace $\mathfrak{H}$ of $L_2(0,2\pi)$ spanned by the elements PV^kf_0, i.e.

$$\mathfrak{H}=\bigvee_{k=0}^{\infty} PV^kf_0.$$

$\mathfrak{H}$ is a non-trivial Hilbert space since it contains in particular the non-zero element f_0.

Since $V^*\mathfrak{M}\subset\mathfrak{M}$, we have $PV^*P=V^*P$, $PVP=PV$, and by induction

$$(PV)^nP=PV^n=PV^nP\quad (n=0,1,\ldots).$$

From these relations it follows that

$$(PV)^nPV^kf_0=PV^nV^kf_0=\begin{cases}PV^{n+k}f_0\\ PV^nPV^kf_0\end{cases}\quad (n,k=0,1,\ldots),$$

thus the restriction of PV to $\mathfrak{H}$ transforms $\mathfrak{H}$ in itself: let it denote by T; we see also that

$$T^ng=PV^ng\quad (n=0,1,\ldots) \tag{3}$$

for all elements of the form $g=PV^kf_0$ $(k\geqslant 0)$, hence for all $g\in\mathfrak{H}$. In particular we have $T^nf_0=PV^nf_0$ $(n\geqslant 0)$, thus

$$\mathfrak{H}=\bigvee_{n=0}^{\infty} T^nf_0. \tag{4}$$

T is obviously a contraction operator, and by (3) it has the (not necessarily minimal) unitary dilation V with absolutely continuous spectral measure. Thus $T\in\tilde{D}$, and by (3)

$$\varphi(T)g=P\varphi(V)g.$$

for all $\varphi(z)\in H_\infty$ and all $g\in\mathfrak{H}$. In particular we have

$$T^n\varphi(T)f_0=PV^n\varphi(V)f_0\quad (n=0,1,\ldots),$$

hence, for $n, m \geqslant 0$,

$$(\varphi(T)\, T^n f_0, T^m f_0) = (\varphi(V)\, V^n f_0, P V^m f_0).$$

The last expression is equal to 0 since $\varphi(V)\, V^n f_0$ is orthogonal to $\mathfrak{M}$. The elements $T^n f_0 (n \geqslant 0)$ span $\mathfrak{H}$, thus it results that

$$\varphi(T) = 0$$

i.e. T is a zero in $\tilde{D}$ of the extended function φ.

Moreover, this T is completely non unitary. For if it were not, there would exist an element $g_0 \in \mathfrak{H}$, $g_0 \neq 0$, such that

$$\|T^n g_0\| = \|g_0\| \quad (n = 0, 1, \ldots),$$

and this would imply

$$T^n g_0 = V^n g_0 \quad (n = 0, 1, \ldots)$$

and hence

$$\varphi(T) g_0 = \varphi(V) g_0,$$

thus $\varphi(V) g_0 = 0$, i.e.

$$\varphi(e^{i\theta}) g_0(\theta) = 0 \quad \text{a.e.}$$

Now, this is impossible since $\varphi(e^{i\theta})$ can vanish only on a set of 0 measure This finishes the proof of (III).

REFERENCES

[1]. Beurling, A., On two problems concerning linear transformations in Hilbert space. *Acta Math.*, 91 (1949), 239–255.

[2]. Sz.-Nagy, B. & Foiaş, C., Sur les contractions de l'espace de Hilbert. IV. *Acta Sci. Math.*, 21 (1960), 251–259.

[3]. —— Sur les contractions de l'espace de Hilbert. VI. Calcul fonctionnel. *Acta Sci. Math.*, 23 (1962), 130–167.

MAXIMAL IDEAL SPACES

By JOHN WERMER

1. Introduction

Let Γ denote the unit circle in the z-plane and D the closed unit disk. Let A_0 be the class of continuous complex-valued functions on Γ which can be extended to D to be analytic in $D-\Gamma$ and continuous in D. A_0 now has the following properties: it is an algebra over C, the complex numbers. It is closed under uniform convergence on Γ. It contains the constants and separates points on Γ.

Each point a in D induces a non-zero homomorphism of A_0 into C, i.e. the map: $f \to f(a)$. Conversely, one can show that every non-zero homomorphism arises from some point of D. Moreover, the maximum principle gives that if f is in A_0 and a in D,

$$|f(a)| \leqslant \max_{z \in \Gamma} |f(z)|. \tag{1}$$

Finally, the Poisson integral formula states that

$$f(a) = \int_\Gamma f(z)\, K(a, z)\, d\theta \tag{2}$$

for a certain positive kernel K.

Let us now replace Γ by an arbitrary compact Hausdorff space X and A_0 by any class A of continuous complex-valued functions on X which satisfies the following:

(a) A is an algebra over C;

(b) A is closed under uniform convergence on X;

(c) A contains the constants and separates points on X.

We call A a *function algebra on* X. In view of what we said above, a reasonable analogue of D here is the space of all non-zero homomorphisms of A into C.

With $\|f\| = \max_X |f|$, A is a Banach algebra. It is evidently commutative, semi-simple and has a unit. Let M denote the space of all non-zero homomorphisms of A into C. The general theory of Banach algebras now gives the following information:

Every maximal ideal of A is the kernel of an element of M, and conversely. For this reason M is called the *maximal ideal space* of A. If f is in A, define the function $\hat{f}$ on M by $\hat{f}(m) = m(f)$. Then M can be topologized so as to become a compact Hausdorff space and such that all the functions $\hat{f}$ are continuous.

Further, each x in X induces a homomorphism m_x in M, namely evaluation at x. By identifying m_x with x, we may regard X as a closed subset of M and $\hat{f}$ as a continuous extension of f from X to all of M. In the case of A_0, $\hat{f}$ is seen to be the analytic extension of f to D.

On a Banach algebra a homomorphism into C is bounded as a linear functional and has norm 1. Hence for f in A and m in M, $|\hat{f}(m)| = |m(f)| \leqslant \|m\| \cdot \|f\| = \|f\| = \max_X |f|$. Thus

$$|\hat{f}(m)| \leqslant \max_X |f|, \tag{1'}$$

i.e. the analogue of the maximum principle holds. Also, since m is a bounded linear functional on A, the Riesz representation theorem gives the existence of a complex measure μ_m on X of total mass 1 with

$$\hat{f}(m) = \int f d\mu_m \quad \text{for all} \quad f \quad \text{in} \quad A. \tag{2'}$$

Since also $\int d\mu_m = 1$, μ_m is real and $\geqslant 0$, [1]. (2′) is thus an analogue of the Poisson formula. Any non-negative measure on X satisfying (2′) will be called a *representing measure for* m.

A number of questions now present themselves. What is the geometric structure of M? In the case of A_0 we have that $M - X = D - \Gamma$ has a complex analytic structure. To what extent does this hold in the general case? What results of the theory of functions, beyond (1′) and (2′), are valid for the functions $\hat{f}$ on $M - X$?

A beginning has been made on the study of these questions during the last 10 years.

It should be pointed out here that for certain algebras $M - X$ is empty, i.e. M coincides with X. This is always true, for instance, if A is the algebra $C(X)$ of *all* continuous functions on X.

2. Finitely generated algebras

We say that A is *generated by* $g_1, \ldots, g_n$ if the polynomials $p(g_1, \ldots, g_n)$ form a dense subalgebra of A.

Let Y be a compact subset of the space C^n of n complex variables. We denote by $P(Y)$ the class of all functions on Y which are uniform limits of polynomials in the coordinates on Y. Evidently, $P(Y)$ is a function algebra on Y generated by n functions. We now have

THEOREM. *Let A be an n-tuply-generated function algebra on a space X. Then X can be mapped homeomorphically on a compact subset X' of C^n in such a way that A goes over into $P(X')$.*

Our initial example A_0 is generated by the function z on the circle Γ and so $A_0 = P(\Gamma)$.

Finitely generated function algebras are important for two reasons. First, their study is identical with the study of uniform approximation by polynomials in several complex variables on compact sets. Second, arbitrary function algebras can be approached through their finitely generated subalgebras.

Let X now be any compact set in C^n. We define $h(X)$, the *hull of* X, as the set of points y in C^n such that for all polynomials P we have

$$|P(y)| \leqslant \max_X |P|. \tag{3}$$

If y is a point in $h(X)$, evaluation at y is a homomorphism of the algebra of polynomials into the complex numbers which admits a unique extension to a homomorphism of $P(X)$. Starting from this observation we easily see that the maximal ideal space of $P(X)$ can be identified with $h(X)$.

A basic result in the study of finitely generated algebras is:

THEOREM. *Let X be a compact set in C^n and F a function analytic in a neighborhood of $h(X)$. Then F, restricted to X, lies in $P(X)$.*

This result is due to Silov [2], and is a consequence of a classical approximation theorem of A. Weil. Stronger forms of this result are given in [3] and [4].

3. Topological properties of M

The following theorems were proved for arbitrary commutative Banach algebras. We state them only for a function algebra A on a compact space X, with maximal ideal space M.

THEOREM (Silov [2]). *M is disconnected if and only if A contains non-trivial idempotents. In particular, X connected implies M connected.*

This is proved from the preceding result by reduction to the finitely generated case.

Let now $H^i(M,G)$ denote the ith Cech cohomology group of M with coefficients in the group G. Arens in [5] and Royden in [6] have proved

THEOREM. *Let U denote the group of units in A (elements with inverses in A) and let U_0 denote the group of elements of the form e^a with a in A. Then $H^1(M,Z)$ is isomorphic to the quotient group U/U_0. (Here Z is the integers.)*

This result had previously been known when $A=C(X)$: (Bruschlinsky [7]). Its proof depends on the theory of analytic sheaves, as developed in the Cartan seminar [8], and again on reduction to the finitely generated case.

When A is finitely generated, more can be said.

THEOREM (A. Browder [9]). *If A is generated by n elements, then $H^i(M,C)=0$ for $i \geqslant n$. (Here C is the complex numbers.)*

This is based on a theorem of Serre [10].

4. The local maximum modulus principle

As we saw, an analogue of the maximum modulus principle of function theory holds in the form (1′) for any function algebra. However, in function theory one also has a local form of this principle which for the disk D may be stated as follows: Let F belong to A_0 and let z_0 be a point of $D-\Gamma$. Let V be any neighborhood of z_0. Then $|F(z_0)| \leqslant$ the maximum of $|F|$ taken over the boundary of V.

A little reflection will show that the generalisation of this form of the maximum principle to function algebras is far from obvious. It is, however, possible and the following holds:

THEOREM (H. Rossi [11]). *Let A be a function algebra on a space X, M its maximal ideal space and x_0 a point of $M-X$. Let V be a neighborhood of x_0 in $M-X$ and ∂V its boundary. Then*

$$|\hat{f}(x_0)| \leqslant \max_{\partial V} |f|, \quad \textit{all} \quad f \quad \textit{in} \quad A. \tag{4}$$

This theorem contains the following geometrical statement about the hull $h(X)$ of a set X in C^n: Let $z_1,\ldots,z_n$ be the coordinates and let S be the hypersurface $|z_1|=1$. Let q be a point of $h(X)$ which lies on S and assume that in some neighborhood of q $h(X)$ lies (except for q) entirely on the side $|z_1|<1$. Then q belongs to X. For else (4) applied to z_1 would give $|z_1(q)|<1$, denying the hypothesis that q lies on S.

5. Parts and representing measures

Recall the algebra A_0 and let z_1, z_2 be two points of D. Define $\varrho(z_1,z_2)=\sup |f(z_1)-f(z_2)|$ taken over all f in A_0 with $\|f\|\leqslant 1$. If both z_1 and z_2 lie in $D-\Gamma$, we get $\varrho(z_1,z_2)<2$, in the following way: let φ be a conformal map of D on itself which takes 0 into z_2 and w_1 into z_1. Let f lie in A_0 and $\|f\|\leqslant 1$. Then $g=f(\varphi)$ is in A_0 and $\|g\|\leqslant 1$. Also

$$|f(z_1)-f(z_2)| = |g(w_1)-g(0)| \leqslant 2\cdot|w_1|$$

by Schwarz's Lemma, since $|g-g(0)|\leqslant 2$ on D. Thus $\varrho(z_1,z_2)\leqslant 2\cdot|w_1|$.

If one or both of z_1, z_2 lie on Γ, $\varrho(z_1,z_2)=2$, as is easily seen.

Gleason was led in [12] to generalize this to function algebras. Let A be any function algebra on a space X and M its maximal ideal space. If m_1, m_2 are in M, put $\varrho(m_1,m_2)=\sup|\hat{f}(m_1)-\hat{f}(m_2)|$ taken over all f in A with $\|f\|\leqslant 1$. In other words, $\varrho(m_1,m_2)$ is the distance between m_1 and m_2 in the dual Banach space of A. Gleason showed.

LEMMA. *The relation: "$\varrho(m_1,m_2)<2$" is an equivalence relation in M.*

He called the equivalence classes the *parts* of M. For the algebra A_0 the parts are the open disk $D-\Gamma$ and the individual points of Γ. Although for general function algebras the parts can apparently be very complicated, for certain classes of algebras to be discussed in the next section they have a simple structure. Parts are related to the representing measures which we defined in the Introduction.

LEMMA (Bishop). *If m_1, m_2 lie in different parts, and if μ_1, μ_2 are representing measures for m_1, m_2 on X, then μ_1 and μ_2 are mutually singular.*

6. Dirichlet algebras

A_0 has the following property: every real-valued continuous function on Γ can be uniformly approximated by real parts of functions in A_0. This suggests that we consider function algebras A on a space X satisfying:

(*d*) Re A is dense in $C_R(X)$,

where Re A denotes the space of functions Re f, f in A, and $C_R(X)$ denotes the space of all real continuous functions on X. Following Gleason [12], we

call an algebra A which satisfies (d) a *Dirichlet algebra on* X, because of the role of (d) in the Dirichlet problem. Dirichlet algebras have many interesting properties generalizing known properties of A_0. (For an account and references see [13] and [14], Chapter 4.) Here we shall only be concerned with the parts.

For the rest of this section, let A be a Dirichlet algebra on a space X and M its maximal ideal space.

LEMMA. *Fix m in M. Then there is only one representing measure for m.*

This follows directly from (d).

LEMMA (Gleason [12]). *If m_1, m_2 lie in the same part of M and if μ_1, μ_2 are the representing measures for m_1, m_2 on X, then μ_1 and μ_2 are mutually absolutely continuous and $d\mu_1/d\mu_2$ is bounded.*

THEOREM (Wermer [15]). *Let P be a part of M containing more than one point. Then there exists a one-one continuous map τ of the open disk into M such that τ maps the disk onto P and such that for each f in A the function $\hat{f}(\tau)$ is analytic on the disk.*

Roughly speaking then, a part of M is either a single point or a disk. Each point of X is easily seen to be a (one-point) part. There exist examples where a point of $M-X$ is a one-point part. It is conjectured that if every part is a single point, then $A=C(X)$.

Hoffman [16] has introduced a class of function algebras B on X which instead of (d) satisfy:

(d') The class of functions $\log |f|$ with both f and $1/f$ belonging to B is dense in $C_R(X)$.

He shows in [16] that most of the properties of Dirichlet algebras, and in particular the three results stated in this section, hold for every B satisfying (d'). It is clear that (d) implies (d'): for if f is in A, $\operatorname{Re} f = \log |e^f|$ and e^f as well as e^{-f} are in A. However, there exist important function algebras which satisfy (d') and not (d).

7. Some examples

For each compact set X in C^n the algebra $P(X)$, defined in section 2, is a function algebra on X. We thus have an abundance of examples. Except for very special cases, however, when $n>1$ it is a difficult problem to compute the hull of X (and hence the maximal ideal space of $P(X)$).

Other examples of function algebras and in particular of Dirichlet algebras are given in [13]. Here we shall give a few examples arising from algebras of analytic functions.

Example 1. Let R be an open Riemann surface and R_1 a region on R bounded by a simple closed analytic curve α. We write A_1 for the space of functions analytic and bounded on $R_1 \cup \alpha$. Assume

$$\text{If } p \in R_1,\ f \in A_1, \quad \text{then} \quad |f(p)| \leqslant \max_{\alpha} |f|. \tag{5}$$

Thus (5) is the maximum principle for A_1 relative to α. (5) holds, for instance, if the ideal boundary of R_1 has harmonic measure 0. Assume further

A_1 separates points on $R_1 \cup \alpha$ and for each p in α there is some Q_p in A_1 with $dQ_p \neq 0$ at p. (6)

Let A denote the closure of A_1 on α. Then A is a function algebra on α. Let M be its maximal ideal space. By (5) and (6) we may regard $R_1 \cup \alpha$ as a subset of M. It is of interest to find the remaining points of M.

THEOREM (Wermer [17]). *M is a compact bordered Riemann surface of finite genus whose border is a simple closed curve. R_1 lies conformally embedded in M.*

Hence one has the purely function theoretic consequence:

COROLLARY. (5) *and* (6) *imply that R_1 has finite genus.*

H. Royden (unpublished) has extended this theorem to the case where only (5) is assumed.

The above theorem is proved by regarding A_1 as an algebra of functions analytic on α and establishing the following general result about algebras of functions analytic on the unit circle Γ.

Lemma [17]. *Let B_0 be an algebra of functions analytic on the circle Γ, and let B be the uniform closure of B_0 on Γ. Assume B_0 separates points on the circle and contains the constants, and that for each t on Γ, $\exists\, g_t \in B_0$ with $g_t' \neq 0$ at t. Assume also $B \neq C(\Gamma)$. Then $\exists$ a Riemann surface F and a conformal embedding of an annulus $1-\epsilon < |z| < 1+\epsilon$ in F with the following properties: Γ maps on a simple closed curve γ which bounds a region Ω on F with $\Omega \cup \gamma$ compact.*

Each f in B goes into a function f^* on γ which admits analytic extension to Ω. (7)

Each homomorphism of B into C is of the form: $f \to f^*(p)$ for some p in $\Omega \cup \gamma$. (8)

Example 2. Instead of considering algebras of analytic functions on a circle, consider now an arbitrary compact set K on a Riemann surface R and an algebra $\mathfrak{A}$ of analytic functions on R. Bishop in [18], Theorem 3, has extended the result of the preceding Lemma to this general situation. Assume $\mathfrak{A}$ separates points on R, contains the constants and contains every function on R which is uniformly approximable on every compact set on R by functions in $\mathfrak{A}$. (The assumption of separation of points is made here to simplify the statement of the result. Bishop does not make it.) Let B be the uniform closure of $\mathfrak{A}$ on K.

THEOREM [18]. *There exists a Riemann surface R' and a conformal embedding σ of R in R' with the following property: denote by K^* the union of $\sigma(K)$ and of those components of $R' - \sigma(K)$ which are relatively compact on R'. Then*

Each f in B goes into some f^* on $\sigma(K)$ which admits an analytic extension to $K^* - \sigma(K)$. (9)

Each homomorphism of B into C is of the form: $f \to f^*(p)$ for some p in K^*. (10)

Example 3. Let D_0 denote the open unit disk. Let H^∞ be the class of all bounded analytic functions on D_0. *A priori* H^∞ is an algebra of functions defined on a non-compact set. We can, however, identify it with a function algebra on a certain compact space, by proceeding as follows:

Let L^∞ be the space of bounded measureable functions on the circle $|z|=1$. In its usual norm, L^∞ is a Banach algebra. Write X for the maximal ideal space of L^∞. Then X is a compact Hausdorff space which is totally disconnected. Looked at on X, L^∞ is seen to be the full algebra $C(X)$.

For each f in H^∞, put $f^*(e^{i\theta})=\lim_{r\to 1} f(re^{i\theta})$. Then $f^{*'}$ is in L^∞ by Fatou's theorem. We can therefore regard H^∞ as embedded in L^∞, and hence as embedded in $C(X)$. It is not hard to verify that H^∞ is a closed subalgebra of $C(X)$, separates points on X, and of course contains the constants. Thus H^∞ is a function algebra on X.

Let M denote the maximal ideal space of H^∞. Clearly M contains D_0 since evaluation at any point of D_0 is a homorphism on H^∞. Also clearly M contains X. M is, however, much larger than $D_0 \cup X$ and the structure of M is quite strange and only partially understood. See [14], Chapter 10, for a detailed discussion of this question.

M was first studied by Kakutani and Chevalley (unpublished) in 1942. An outstanding problem until this year was this: Is D_0 dense in M? An equivalent question is:

Let $f_1,\dots,f_n$ be functions in H^∞ such that

$$|f_1(\lambda)|+|f_2(\lambda)|+\dots+|f_n(\lambda)|\geqslant\delta>0 \quad (|\lambda|<1).$$

Do there then exist functions $g_1,\dots,g_n$ in H^∞ such that

$$f_1g_1+f_2g_2+\dots+f_ng_n=1?$$

Carleson [19] has answered this question affirmatively. Earlier partial results on the problem are due to Carleson [20] and Newman [21].

It is easy to see that H^∞ satisfies condition (d') of Section 6. For let u be a real continuous function on X. We consider u as a function in L^∞. Denote by U the bounded harmonic function in D_0 which has u as its boundary function, and put $f=\exp(U+iV)$, where V is the harmonic conjugate of U. Clearly f and $1/f$ belong to H^∞ and on X $\log|f|=u$. Thus (d') holds. Hence the results of Hoffman in [16] apply to H^∞. In particular, all the parts of M are points or disks.

Example 4. Let J be any Jordan arc in the plane. Denote by A_J the algebra of functions which are continuous on the whole Riemann sphere and analytic outside J. For rectifiable arcs, A_J reduces to the constants. When J has positive 2-dimensional Lebesgue measure, it was shown by Denjoy in 1932 that the function

$$f(\zeta)=\int_J \frac{dx\,dy}{z-\zeta}$$

is non-constant and belongs to A_J.

The author pointed out in [22] that whenever A_J does not reduce to the constants, then A_J separates points on the sphere S and, restricted to J, is a function algebra on J. Furthermore, in this case the maximal ideal space of A_J is the sphere S, as was shown by Arens in [23]. In [24] A. Browder and the author showed the following: Let J be an arc and let φ_J be a conformal map of $S-J$ on D_0 under which J corresponds to both the upper and the lower semi-circle on $|z|=1$. Let ψ_J be the induced identification

map of the upper on the lower semi-circle. If we parametrise $|z|=1$ as $z=e^{i\theta}, -\pi \leqslant \theta \leqslant \pi$, ψ_J becomes a continuous decreasing function on $[0, \pi]$ with values in $[0, -\pi]$ and $\psi_J(0)=0$, $\psi_J(\pi)=-\pi$.

THEOREM. *A_J is a Dirichlet algebra on J if and only if ψ_J is singular (i.e. if the derivative $\psi_J'=0$ a.e.)*

This has the purely function theoretic consequence that if J is an arc with ψ_J singular, then A_J contains non-constant functions. Such arcs J exist.

Example 5. Let A be a function algebra on a space X and M its maximal ideal space. One may ask: does $M-X$ possess some kind of analytic structure relative to which the functions $\hat{f}$, f in A, are analytic functions? Stolzenberg [25] has given an example which answers this question in the negative.

REFERENCES

[1]. ARENS, R. & SINGER, I. M., Function values as boundary integrals. *Proc. Amer. Math. Soc.*, 5 (1954), 735–745.

[2]. ŠILOV, G. E., On the decomposition of a commutative normed ring into a direct sum of ideals. *Mat. Sb.*, 32 2 (1953), 353–364.

[3]. ARENS, R. & CALDERÓN, A. P., Analytic functions of several Banach algebra elements. *Ann. Math.*, 62 (1955), 204–216.

[4]. WAELBROECK, L., Le calcul symbolique dans les algèbres commutatives. *J. Math. Pures Appl.*, 33 (1954), 147–186.

[5]. ARENS, R., The group of invertible elements of a commutative Banach algedra. *Proc. Functional Analysis Conf.* Warsaw, 1960.

[6]. ROYDEN, H. L., On the multiplicative groups of function algebras. *Tech. Rep. No. 18*, Stanford Univ. (1960).

[7]. BRUSCHLINSKY, N., Stetige Abbildungen und Bettische Gruppen der Dimensionszahl 1 und 3. *Math. Ann.* 109 (1934), 525–537.

[8]. CARTAN, H., *Séminaires*, Paris, (1951–1952).

[9]. BROWDER, A., Cohomology of maximal ideal spaces. *Bull. Amer. Math. Soc.*, 67 5 (1961), 515–516.

[10]. SERRE, J.-P., Une propriété topologique des domaines de Runge. *Proc. Amer. Math. Soc.*, 6 (1955), 133–134.

[11]. ROSSI, H., The local maximum modulus principle. *Ann. Math.*, 72 1 (1960), 1–11.

[12]. GLEASON, A. M., Function algebras. *Seminars on Analytic Functions*, Vol. II. Institute for Advanced Study, Princeton, 1957.

[13]. WERMER, J., Banach algebras and analytic functions. *Advances in Mathematics*, Vol. 1, Fasc. 1. Academic Press, 1961.

[14]. HOFFMAN, K., *Banach Spaces of Analytic Functions*. Prentice-Hall, 1962.

[15]. WERMER, J., Dirichlet algebras. *Duke Math. J.* 27 (1960) 373–382.

[16]. HOFFMAN, K., Analytic functions and logmodular Banach algebras (Multigraphed, 1962).

[17]. WERMER, J., The maximum principle for bounded functions. *Ann. Math.*, 69, 3 (1959), 598–604.

[18]. BISHOP, E., Analyticity in certain Banach algebras. *Trans. Amer. Math. Soc.*, 102 3 (1962), 507–544.

[19]. CARLESON, L., Interpolations by bounded analytic functions and the Corona problem. *Ann. Math.* 76 (1962), 547–559.

[20]. —— On bounded analytic functions and closure problems. *Ark. Mat.* 2 12 (1952), 283–291.

[21]. NEWMAN, D. J., Some remarks on the maximal ideal structure of H^∞, *Ann. Math.*, 70 3 (1959), 438–445.

[22]. WERMER, J., Polynomial approximation on an arc in C^3. *Ann. Math.*, 62 (1955), 269–270.

[23]. ARENS, R., The maximal ideals of certain function algebras. *Pacific J. Math.*, 8 (1958), 641–648.

[24]. BROWDER, A. & WERMER, J., Some algebras of functions on an arc. *J. Math. Mech.*, 12 (1963), 119–130.

[25]. STOLZENBERG, G., A hull with no analytic structure, *J. Math. Mech.*, 12 (1963), 103–111.

APPLICATIONS OF THE GROTHENDIECK–ATIYAH–HIRZEBRUCH FUNCTOR $K(X)$

By J. F. ADAMS

I shall begin by stating three results; then I shall comment on their authorship and history, and finally I shall try to show how they can be fitted into a general theory.

Let $O(n)$ be the orthogonal group; then we can consider the coset space $O(n)/O(n-k)$ and the fibering

$$O(n)/O(n-k) \to O(n)/O(n-1) = S^{n-1}. \tag{1}$$

The classical problem about vector-fields on spheres used to ask: for what values of n and k does this fibering admit a cross-section? The answer is as follows.

THEOREM 1. *The fibering* (1) *admits a cross-section if and only if n is divisible by N_k, where N_k is the integer defined below.*

We define $N_k = 2^{a(k)}$, where $a(k)$ is the number of integers r such that $1 \leqslant r \leqslant k-1$ and $r \equiv 0$, 1, 2 or 4 mod 8.

Similarly, let $U(n)$ be the unitary group; then we can consider the coset space $U(n)/U(n-k)$ and the fibering

$$U(n)/U(n-k) \to U(n)/U(n-1) = S^{2n-1}. \tag{2}$$

Again we ask: when is there a cross-section?

THEOREM 2. *The fibering* (2) *admits a cross-section if and only if n is divisible by M_k, where M_k is the integer defined below.*

We define $\nu_p(n)$ to be the exponent of the prime p in the decomposition of n into prime powers, so that

$$n = 2^{\nu_2(n)} 3^{\nu_3(n)} 5^{\nu_5(n)} \ldots$$

We define M_k as follows:

$$\nu_p(M_k) = 0 \quad \text{if} \quad p > k,$$

$$\nu_p(M_k) = \operatorname{Sup}\,(r + \nu_p(r)) \quad \text{if} \quad p \leqslant k,$$

where the integer r runs over the range

$$1 \leqslant r \leqslant \frac{k-1}{p-1}.$$

At the last congress we heard a lecture by J. Milnor [10], which was in part about the J-homorphism of H. Hopf and G. W. Whitehead. I recall that this is a map

$$J : \pi_l(SO(n)) \to \pi_{n+l}(S^n).$$

I shall suppose that $l=4k-1$ and $n>4k$, so that we are dealing with the "stable J-homomorphism", and J is defined on a cyclic infinite group. The problem is to describe the image of J. Following Milnor and Kervaire [10], we define $m(k)$ to be the denominator of $B_k/4k$, when this fraction is expressed in its lowest terms. Here B_k is the kth Bernoulli number, so that

$$\frac{x}{e^x-1}=1-\frac{x}{2}+\sum_{k=1}^{\infty}(-1)^{k-1}B_k\frac{x^{2k}}{(2k)!}.$$

THEOREM 3. *If $l=4k-1$ and $n>4k$, then*

$$\operatorname{Im} J=\begin{cases} Z_{m(k)} & \text{if} \quad 4k\equiv 4 \bmod 8, \\ \text{either } Z_{m(k)} \quad \text{or} \quad Z_{2m(k)} & \text{if} \quad 4k\equiv 0 \bmod 8. \end{cases}$$

In Theorem 1, the existence of a cross-section when n is divisible by N_k is classical; the non-existence of a cross-section when n is not divisible by N_k is proved in [1].

The two halves of Theorem 2 were obtained in the reverse order. The non-existence of a cross-section when n is not divisible by M_k is due to Atiyah and Todd [7]; the existence of a cross-section when n is divisible by M_k is proved in [3].

In Theorem 3, the fact that the order of Im J is a multiple of $m(k)$ is the result of Milnor and Kervaire [10] as improved by Atiyah and Hirzebruch [5]. The fact that the order of Im J divides $m(k)$or $2m(k)$, as the case may be, is proved in [2].

I remark that by re-proving the result of Milnor and Kervaire, one can extract more information. One can show [2] that there is a homomorphism

$$e:\pi_{n+4k-1}(S^n)\to Z_{m(k)},$$

such that the composite eJ is an epimorphism (if $n>4k$). It follows that if $4k\equiv 4 \bmod 8$, then Im J is a direct summand in $\pi_{n+4k-1}(S^n)$. If $4k\equiv O \bmod 8$, then a similar conclusion follows except for the 2-component. The homomorphism e has other interesting properties, on which I shall not dwell.

In order to prove these theorems one makes use of the "extraordinary cohomology theory" $K(X)$ of Grothendieck–Atiyah–Hirzebruch [5,6]. I will now recall how this is constructed. Let Λ denote either the real field R or the complex field C. Let X be a "good" space, e.g. a finite connected CW-complex. If $\Lambda=R$ we take all orthogonal bundles over X; if $\Lambda=C$ we take all unitary bundles over X. In either case we divide them into isomorphism classes $\{\xi\}$. We take these classes as the generators for a free abelian group $F_\Lambda(X)$. We shall define

$$K_\Lambda(X)=F_\Lambda(X)/T_\Lambda(X),$$

so that $K_\Lambda(X)$ is given by the generators $\{\xi\}$ and certain relations; we define $T_\Lambda(X)$ to be the subgroup of $F_\Lambda(X)$ generated by all elements of the form

$$\{\xi\oplus\eta\}-\{\xi\}-\{\eta\},$$

where $\xi\oplus\eta$ denotes the Whitney sum of ξ and η. The group $K_\Lambda(X)$, then, is obtained by taking the vector bundles over X and forcing them to generate

an abelian group under the Whitney sum operation. The elements of $K_\Lambda(X)$ may be called "virtual bundles".

It is possible to use the groups $K_\Lambda(X)$ to prove non-existence results in just the same way that one is accustomed to use ordinary cohomology groups. Thus, if X, A is a pair and A is a retract of X, it follows that $K_\Lambda(A)$ is a direct summand in $K_\Lambda(X)$; and if we find that $K_\Lambda(A)$ is not a direct summand in $K_\Lambda(X)$, then we can conclude that A is not a retract of X. The non-existence proof in [1] is presented in this way.

However, just as in ordinary cohomology we often need to use cohomology operations, so here we need to use cohomology operations in $K_\Lambda(X)$.

The first such operation is a cup-product. We can define the tensor product of two vector-spaces over Λ; therefore we can define the tensor product $\xi \otimes \eta$ of two vector bundles over X; one shows that this defines a product in $K_\Lambda(X)$.

Similarly, we can define the dual of a vector-space over Λ; therefore we can define the dual ξ^* of a vector bundle over X; one shows that this defines an operation in $K_\Lambda(X)$.

Again, we can define the ith exterior power of a vector space over Λ; therefore we can define the ith exterior power $\lambda^i(\xi)$ of a bundle over X. It is possible to extend the definition of λ^i from bundles to virtual bundles in a unique way so as to preserve the following familiar property:

$$\lambda^i(\xi+\eta) = \sum_{j+k=i} \lambda^j(\xi) \otimes \lambda^k(\eta).$$

All this is due to Grothendieck.

Unfortunately, the formal properties of the λ^i are not very convenient. It is possible to obtain operations with better formal properties by an algebraic device. Consider

$$(x_1)^k + (x_2)^k + \ldots + (x_k)^k;$$

this is a symmetric polynomial in $x_1, x_2, \ldots, x_k$; therefore it can be written as a polynomial in the elementary symmetric functions σ_i of $x_1, x_2, \ldots, x_k$; say

$$(x_1)^k + (x_2)^k + \ldots + (x_k)^k = Q^k(\sigma_1, \sigma_2, \ldots, \sigma_k).$$

Now define

$$\Psi^k(\xi) = \begin{cases} Q^k(\lambda^1\xi, \lambda^2\xi, \ldots, \lambda^k\xi), & (k>0), \\ \Psi^{-k}(\xi^*), & (k<0). \end{cases}$$

The functions Ψ are ring homomorphisms from $K_\Lambda(X)$ to $K_\Lambda(X)$.

In order to unify the three theorems with which I started, one makes use of the groups $J(X)$ of Atiyah [4]. First I define the notion of fibre homotopy equivalence. Let ξ, η be sphere bundles over X, with total spaces E_ξ, E_η; then we say that a map $f: E_\xi \to E_\eta$ is "fibrewise" if it covers the identity map of X; we say that ξ, η are fibre homotopy equivalent if there exist fibrewise maps $f: E_\xi \to E_\eta$, $g: E_\eta \to E_\xi$ such that $gf \sim 1$ through fibrewise maps of E_ξ, and similarly for fg. We shall define

$$J_\Lambda(X) = K_\Lambda(X)/U_\Lambda(X),$$

so that $J_\Lambda(X)$ is given by the generators $\{\xi\}$ and certain relations; we

define $U_\Lambda(X)$ to be the subgroup of $K_\Lambda(X)$ generated by all elements of the form

$$\{\xi\}-\{\eta\},$$

where ξ,η are fibre homotopy equivalent.

Since all our groups are functorial, we can write

$$K_\Lambda(X)=K_\Lambda(P)+\tilde{K}_\Lambda(X),$$

$$J_\Lambda(X)=J_\Lambda(P)+\tilde{J}_\Lambda(X);$$

here P denotes a point, and these equations are supposed to define the summands $\tilde{K}_\Lambda(X)$, $\tilde{J}_\Lambda(X)$ complementary to $K_\Lambda(P)$, $J_\Lambda(P)$. Atiyah's group $J(X)$ is the one I have called $\tilde{J}_R(X)$.

According to Atiyah [4], if you can compute $J_R(RP^{k-1})$, you can prove Theorem 1; if you can compute $J_C(CP^{k-1})$, you can prove Theorem 2; and we have

$$J_R(S^l)=J(\pi_{l-1}(SO(n)))\quad (n>l).$$

We therefore face the general problem: "compute $J_\Lambda(X)$".

Half of the problem consists in giving a lower bound for $J_\Lambda(X)$, and half of it consists in giving an upper bound for $J_\Lambda(X)$. I start with the lower bound.

It is sometimes easy to prove that two bundles ξ,η are not fibre homotopy equivalent by using the Stiefel-Whitney classes, which are fibre homotopy invariants. The reason why they are fibre homotopy invariants is that they can be defined in a particular way. Suppose given a sphere bundle ξ over B with total space E; we can embed the space E in the corresponding bundle of unit solid balls, say $\bar{E}$. Then in cohomology we have the Thom isomorphism

$$\varphi_H: H^*(B;Z_2)\to H^*(\bar{E},E;Z_2).$$

We can consider the following diagram.

$$\begin{array}{ccc} & Sq=\sum_{i=0}^{\infty} Sq^i & \\ H^*(\bar{E},E;Z_2) & \longrightarrow & H^*(\bar{E},E;Z_2) \\ \varphi_H\uparrow & & \uparrow\varphi_H \\ H^*(B;Z_2) & & H^*(B;Z_2) \end{array}$$

The total Stiefel-Witney class is given by

$$w(\xi)=\varphi_H^{-1}Sq\,\varphi_H 1$$

We can copy this procedure using the K-cohomology theory. For example, suppose that ξ is a unitary bundle, and let the other notation remain as before. The one can define a Thom isomorphism

$$\varphi_K: K_C(B)\to K_C(\bar{E},E),$$

where the relative K groups are defined by

$$K_\Lambda(X,Y)=\tilde{K}_\Lambda(X/Y).$$

One can consider the following diagram.

$$\begin{array}{ccc} K_C(\bar{E}, E) & \xrightarrow{\Psi^k} & K_C(\bar{E}, E) \\ \varphi_K \uparrow & & \uparrow \varphi_K \\ K_C(\mathrm{B}) & & K_C(B) \end{array}$$

By analogy, we define $\varrho^k(\xi) = \varphi_K^{-1}\Psi^k\varphi_k\, 1$.

We next seek to extend the definition of ϱ^k to virtual bundles, so as to preserve the following property.

$$\varrho^k(\xi+\eta) = \varrho^k(\xi)\varrho^k(\eta).$$

The extension is possible and unique, provided we intepret $\varrho^k(\xi)$ as an element of the group

$$K_C(B) \otimes Q_k,$$

where Q_k denotes the additive group of fractions a/k^b. It is easy to see that we are forced to introduce these denominators. In fact, we have

$$\varrho_k(1) = k,$$

so

$$\varrho_k(-1) = 1/k.$$

For completeness I add that one can also adopt an intermediate approach, and consider (for example) the following diagram.

$$\begin{array}{ccc} K_C(\bar{E}, E) & \xrightarrow{ch} & H^*(\bar{E}, E; Q) \\ \varphi_K \uparrow & & \uparrow \varphi_H \\ K_C(B) & & H^*(B; Q) \end{array}$$

This method yields criteria which can be stated in terms of characteristic classes. However, it is not likely to be adequate if B has torsion; and it also fails to give best possible results for such torsion-free spaces as S^{8m+2}, CP^{4m+1}.

I therefore adopt the following definition of a quotient group $J'_\Lambda(X)$ of $J_\Lambda(X)$, which will serve as a lower bound for $J_\Lambda(X)$. I define

$$J'_\Lambda(X) = K_\Lambda(X)/V,$$

where $x \in V$ if and only if there exists y in $\tilde{K}_\Lambda(X)$ such that

$$\varrho^k(x) = \frac{\Psi^k(1+y)}{1+y} \quad \text{for all } k.$$

(The experts will understand that in the case $\Lambda = R$ we impose also the conditions $w_1(x) = 0$ and $w_2(x) = 0$, in order that $\varrho^k(x)$ should be defined. Compare [8].)

The reason for adopting a definition of this form is that when one tries to prove that $\varrho^k(x)$ is a fibre homotopy invariant, one finds only that it is an invariant up to multiplication by a factor of the form

$$\frac{\Psi^k(1+y)}{1+y}.$$

I will now pass on to discuss upper bounds for $J_\Lambda(x)$. For this purpose we need a result which will prove that two sphere bundles ξ and η are fibre homotopy equivalent, although they are not isomorphic. I offer the following.

THEOREM 4. *Suppose that ξ,η are sphere bundles over a finite CW-complex and that there is a fibrewise map*

$$f: E_\xi \to E_\eta$$

of degree k on each fibre. Then there exists an integer e such that the Whitney multiples $|k^e|\xi$, $|k^e|\eta$ are fibre homotopy equivalent.

If we put $k=1$ this is a theorem of Dold [9]. Therefore one may regard this theorem as a mod k analogue of Dold's theorem. The proof of Dold may be summarised by saying that we consider the space of homotopy equivalences from S^{n-1} to S^{n-1}, and treat it seriously as a "structural group". My proof may be summarised by saying that we take the space of all maps from S^{n-1} to S^{n-1}, and treat it similarly.

By applying Theorem 4, I prove the following.

THEOREM 5. *Suppose that k is given, that $y \in K_C(X)$ (where X is a finite CW-complex) and either* (i) *y is a linear combination of complex line bundles over X, or* (ii) *$X = S^{2n}$.*
Then there exists an integer $e = e(k,y)$ such that

$$k^e(\Psi^k - 1)y$$

maps to zero in $J_C(X)$.

Following the hint contained in Theorem 5, I define a group $J''_\Lambda(X)$ which will act as a sort of conditional upper bound for X; that is to say, if the conclusion of Theorem 5 holds for every pair (k,y), then $J_\Lambda(X)$ will be a homomorphic image of $J''_\Lambda(X)$, no matter how large the integers $e(k,y)$ turn out to be. I define

$$J''_\Lambda(X) = K_\Lambda(X)/W,$$

where $x \in W$ if and only if for every function $e(k,y)$ there exists a function $a(k,y)$, such that

$$x = \sum_{k,y} a(k,y)\, k^{e(k,y)}(\Psi^k - 1)y.$$

It is understood that the functions $e(k,y)$ and $a(k,y)$ are defined for all pairs consisting of an integer k and an element $y \in K_C(X)$; the values of $e(k,y)$ are non-negative integers; $a(k,y)$ takes integer values, and is zero except for a finite number of pairs (k,y).

If we hope to estimate $J_\Lambda(X)$ by means of an upper bound and a lower bound, it is desirable to have these two bounds close together. If $X = RP^{k-1}$, CP^{k-1} or S^l, then

$$J''_R(X) = J'_R(X);$$

and this completes what I want to say about Theorems 1, 2, 3.

It would appear that we have

$$J''_R(X) = J'_R(X)$$

for any finite CW-complex X.

Problem 1. Does the conclusion of Theorem 5 hold for each finite CW-complex X and each element y in $K_R(X)$?

Problem 2. Can Theorem 4 be used to answer Problem 1?

Problem 3. Suppose given two inequivalent representations of $O(n)$, so that it acts on Euclidean spaces V, V' of the same dimension. When can one find an equivariant map $f: V \to V'$ which maps the unit sphere of V onto that of V' with degree k?

REFERENCES

[1]. ADAMS, J. F., Vector fields on spheres. *Ann. Math.*, 75 (1962), 603–632.

[2]. —— On the J-homomorphism. (To appear.)

[3]. ADAMS, J. F. & WALKER, G., On complex Stiefel manifolds. (To appear.)

[4]. ATIYAH, M. F., Thom complexes. *Proc. London Math. Soc.* (*3*), 11 (1961), 291–310.

[5]. ATIYAH, M. F. & HIRZEBRUCH, F., Riemann-Roch theorems for differentiable manifolds. *Bull. Amer. Math. Soc.*, 65 (1959), 276–281.

[6]. ATIYAH, M. F., & HIRZEBRUCH, F., Vector bundles and homogeneous spaces. *Proc. Symp. Pure Maths. 3. Differential Geometry*, 7–38. Amer. Math. Soc., 1961.

[7]. ATIYAH, M. F. & TODD, J. A., On complex Stiefel manifolds. *Proc. Cambridge Philos. Soc.*, 56 (1960), 342–353.

[8]. BOTT, R., A note on the KO-theory of sphere bundles. *Bull. Amer. Math. Soc.*, 68 (1962), 395–400.

[9]. DOLD, A., Über fasernweise Homotopieäquivalenz von Faserräume. *Math. Z.*, 62 (1955), 111–136.

[10]. MILNOR, J. & KERVAIRE, M. A., Bernoulli numbers, homotopy groups, and a theorem of Rohlin. *Proc. Internat. Congress Math. Edinburgh*, 1958, 454–458. Cambridge Univ. Press, 1960.

THE GROTHENDIECK RING IN GEOMETRY AND TOPOLOGY

By M. F. ATIYAH

§ 1. The Grothendieck ring in homotopy theory

I am going to be talking about vector bundles, i.e. fibre bundles with fibre a vector space and group the linear group. Vector bundles are to the geometer what representations or modules are to the algebraist. In fact the modern algebraic geometer hardly distinguishes between the two. Now, in the study of groups, representation theory is important for two reasons: (*a*) because representations of groups occur naturally in many contexts, and (*b*) because characters provide a powerful technical tool with which to attack groups.

The same essentially holds for vector bundles. That they occur naturally in many geometrical problems is of course well known. It is less obvious, but equally true, that, when formalized into the Grothendieck ring, they provide a powerful tool in homotopy theory. The Grothendieck ring is a very simple object which can be defined in a wide range of categories. In the classical representation theory of finite groups one obtains the ring of generalized characters (whose elements are integral linear combinations of the irreducible characters). For vector bundles the definition is as follows.

Let X be a "reasonable" space (say a polyhedron) and consider complex(1) vector bundles over X. If E, F are two vector bundles then we can form their direct sum $E \oplus F$ and their tensor product $E \otimes F$. The operation $\oplus$ turns the set $\mathcal{E}(X)$ of isomorphism classes of vector bundles into an abelian semi-group. $K(X)$ is defined to be the associated abelian group. This means that $K(X)$ is an abelian group with a homomorphism θ: $\mathcal{E}(X) \to K(X)$ such that

(i) Im (θ) generates $K(X)$,

(ii) $\theta(E) = \theta(F)$ if and only if $\phi(E) = \phi(F)$ for every homomorphism ϕ of $\mathcal{E}(X)$ into an abelian group.

It is elementary to show that $K(X)$, θ exist and that (i), (ii) make them essentially unique. The operation $\otimes$ then turns $K(X)$ into a commutative ring—the Grothendieck ring. It is a contravariant functor of X.

Now character theory is essentially a comparison of a given group with the linear groups, and depends for its success on our detailed knowledge of the linear groups. In the same way the study of $K(X)$ is essentially a comparison, in the sense of homotopy, of X with the linear groups (or rather their classifying spaces, the Grassmannians), and depends for its success on detailed knowledge of the homotopy of the linear groups. This essential information is provided by Bott's periodicity theorem which, in our formulation, asserts [5];

$$K(X \times S^2) \cong K(X) \otimes_Z K(S^2), \tag{1}$$

(1) As for representations the complex case is simpler but the real theory can also be developed.

where S^2 is the 2-sphere. This basic formula really states that the (stable) homotopy of the complex linear group is just what linear algebra dictates. For this reason the general machinery which one can develop from (1) is very simple and enables K to be handled with great ease. For instance the computation of $K(X)$ for many homogeneous spaces turns out to be a fairly easy matter (cf. [2]).

The formal properties of K are similar in many ways to those of H (cohomology) and it can be applied in a similar way in problems of homotopy theory. Cohomology operations ($H \to H$) can be replaced either by characteristic classes ($K \to H$) or by operations ($K \to K$). It was pointed out by Grothendieck that one can very easily define operations ($K \to K$) by using operations on vector spaces, and this has been exploited by Adams(1) [1] in his solution of the long-outstanding problem of vector-fields on spheres.

I do not propose to discuss these questions in any further detail, but I will just make a few general remarks about the method. In orthodox algebraic topology one breaks a space down into atomic parts (cohomology or homotopy) and then tries to see how the atoms were put together (operations or k-invariants). Now many of the spaces that turn up in practice, e.g. homogeneous spaces, have such a regular structure that it is a pity to break them right down. Instead, like the biochemist, we should look around for large standard molecules out of which these spaces are built. It would seem that the linear groups provide such standard molecules and that this, in a philosophical sense, explains the success of K in homotopy theory.

§ 2. Geometric applications

Because vector bundles arise so naturally it is to be expected that K, besides its general use in homotopy theory, will have special relevance in more geometric contexts. In order to describe some examples I must now explain one technical point.

If Y is a subspace of X then one can define a relative group $K(X, Y)$ by considering vector bundles on X with a given trivialization on Y. Suppose now that E, F are two vector bundles on X and that $\alpha: E \to F$ is an isomorphism on Y. Then by adding to both E and F a "complement" of F (i.e. an F' so that $F \oplus F'$ is trivial) we get an element

$$d(E, F, \alpha) \in K(X, Y)$$

called the "difference bundle", and this is independent of the choice of F'. I shall now discuss three examples in which this "difference bundle" arises naturally and leads to interesting results.

§ 2.1. Sub-manifolds

Let X be a compact oriented n-manifold embedded in S^{n+2k}. Choose a metric, let N be a tubular neighbourhood of X and put $N_0 = N - X$. The normal bundle ν has structure group $SO(2k)$ but when lifted up to N_0 this

(1) Actually Adams uses K for real vector bundles. Also his operations are a modification of those introduced by Grothendieck.

reduces to $SO(2k-1)$. Now let V be any representation space of $SO(2k)$, and let σ be the outer automorphism of $SO(2k)$ corresponding to the reflection in the plane $x_{2k}=0$. Then V^σ will be a new representation of $SO(2k)$ but, since σ is the identity on $SO(2k-1)$, V and V^σ coincide as representations of $SO(2k-1)$. Take the vector bundles E, E^σ associated to ν (lifted up to N) by the representations V, V^σ. Then on N_0 we have a canonical isomorphism $\alpha: E \to E^\sigma$ and hence a difference bundle

$$d(E, E^\sigma, \alpha) \in K(N, N_0).$$

Using the natural map

$$K(N, N_0) \to K(S^{n+2k})$$

we end up with an element of $K(S^{n+2k})$. But we know all about this group; in particular we know all about characteristic classes on S^{n+2k}. From this we deduce information about the characteristic numbers of X. These results are of two types:

(i) (stable) since X can always be embedded in S^{n+2k} for large k we get some results for any X; these are of importance for example in the study of differentiable structures on manifolds;

(ii) (unstable) taking a low value of k may, for given X, give results which are false; this shows that X cannot be embedded in S^{n+2k} (or $\mathbf{R}^{n+2k}$).

An example of (ii) is the following result [3]:
Complex projective n-space cannot be embedded differentiably in $\mathbf{R}^{4n-2\alpha(n)}$, *where* $\alpha(n)$ *is the number of* 1's *in the dyadic expansion of* n.

§ 2.2. Sheaves

Let X be a complex manifold, S a coherent analytic sheaf on X and suppose we have an exact sequence of coherent sheaves:

$$0 \to E \xrightarrow{\alpha} F \to S \to 0, \tag{2}$$

where E, F are locally free. Then E, F are germs of holomorphic sections of vector bundles (still denoted by E, F) and α gives a vector bundle homomorphism (still denoted by α). Outside $|S| = \text{support}(S)$ α is an isomorphism, and hence we get a difference bundle,

$$d(E, F, \alpha) \in K(X, X - |S|). \tag{3}$$

More generally if, instead of (2), we have a (finite) resolution $\mathcal{L}$ of S by locally free sheaves then we can generalize (3) and define an element

$$d(\mathcal{L}) \in K(X, X - |S|), \tag{4}$$

which turns out to depend only on S and not on $\mathcal{L}$. We can apply this in particular when S is the sheaf of a complex analytic subvariety of X. Using

the element given by (4) and the relations between K and H one can prove the following [4]:(1)

If $u \in H^{2k}(X; \mathbf{Z})$ is represented by a complex analytic subvariety (of codimension k) then $Sq^3 u = 0$.

Here Sq^3 is the Steenrod operation

$$H^{2k}(X; \mathbf{Z}) \to H^{2k+3}(X; \mathbf{Z}).$$

This result is of interest because it disproves an old conjecture of Hodge.

§ 2.3. Elliptic operators

Let X be a compact oriented C^∞-manifold, E, F two C^∞ n-dimensional complex vector bundles over X, $\Gamma(E)$ and $\Gamma(F)$ the spaces of C^∞ cross-sections. Then by a partial differential operator from E to F one means a linear operator

$$D: \Gamma(E) \to \Gamma(F),$$

which in local coordinates is expressed by a matrix of partial derivatives. If we consider only the highest(2) order derivatives and regard $\partial/\partial x_i$ not as an operator but as a coordinate in the bundle T of cotangent vectors of X, we get a well-defined homomorphism

$$\sigma(D): \pi^*(E) \to \pi^*(F),$$

where $\pi: T \to X$ is the bundle projection. This is called the symbol of the operator D, and D is said to be *elliptic* if $\sigma(D)$ is an isomorphism on T_0, the set of non-zero tangent vectors. Hence an elliptic operator defines a difference bundle

$$\gamma(D) = d(\pi^* E, \pi^* F, \sigma(D)) \in K(T, T_0).$$

It appears that the interesting invariants of D are actually invariants of $\gamma(D)$. More specifically D has an invariant called the index defined by

$$\nu(D) = \dim \operatorname{Ker} D - \dim \operatorname{Coker} D$$

and it seems reasonable to conjecture that $\nu(D)$ is expressible by a universal polynomial in terms of the characteristic classes of $\gamma(D)$ and those of X. This is closely related to the Riemann–Roch theorem.

I should just end by saying that most of this is joint work with F. Hirzebruch, while the remarks in § 2.3 are based on some current work, not yet completed[3], with I. M. Singer.

---

(1) This also follows, if X is algebraic, from the resolution of singularities (Hironaka). In fact one then has $Sq^r u = 0$ for all odd r.

(2) Precisely we take derivatives of order k if, somewhere on X, D involves derivatives of order k, but nowhere does D involve derivatives of order $k+1$.

(3) (Added in proof.) See a forthcoming note in the Bulletin of the American Mathematical Society.

REFERENCES

[1]. ADAMS, J. F., Vector fields on spheres. *Ann. Math.*, 75 (1962), 603–632.

[2]. ATIYAH, M. F. & HIRZEBRUCH, F., Vector bundles and homogeneous spaces. *Amer. Math. Soc. Symp.*, III (1961), 7–38.

[3]. —— Théorèmes de non-plongement pour les variétés différentiables. *Bull. Soc. Math. France*, 87 (1959), 383–396.

[4]. —— Analytic cycles on complex manifolds. *Topology*, 1 (1962), 25–46.

[5]. BOTT, R., Quelques remarques sur les théorèmes de périodicité. *Bull. Soc. Math. France*, 87 (1959), 293–310.

LES VARIÉTÉS RIEMANNIENNES DONT LA COURBURE SATISFAIT CERTAINES CONDITIONS

Par M. BERGER

1. Variétés riemanniennes

Dans toute la suite V désigne une variété indéfiniment différentiable, de dimension d supérieure ou égale à deux. L'espace fibré de ses vecteurs est noté $T(V)=\bigcup_{m\in V}T(V)_m$ et $p:T(V)\to V$ est la projection canonique. $F^\infty(V)$ désigne l'anneau des fonctions réelles indéfiniment différentiables de V et $\mathfrak{X}(V)$ le module sur $F^\infty(V)$ constitué par les champs de vecteurs indéfiniment différentiables de V. Une *structure riemannienne* (s.r.) sur V est la donnée d'une forme bilinéaire symétrique $\langle , \rangle$ sur le module $\mathfrak{X}(V)$ qui, si l'on désigne encore par $\langle , \rangle$ et par abus de langage la forme bilinéaire induite sur chaque $T(V)_m$ par $\langle , \rangle$, vérifie la condition $\langle x,x\rangle>0$ quel que soit $x\in T(V)_m$, $x\neq 0$. On en déduit une norme $\|x\|=(\langle x,x\rangle)^{\frac{1}{2}}$. On dit aussi que V est une variété riemannienne (v.r.). Une *isométrie* $f:V\to W$ entre deux v.r. V,W est, par définition, un difféomorphisme f tel que $\|(df)(x)\|=\|x\|$ quel que soit $x\in T(V)$.

2. Exemples

(A) Soit R^d l'espace vectoriel canonique formé avec les d-uples de nombres réels, muni de sa structure euclidienne canonique v.w définie par $(a_1,\dots,a_d)\cdot(b_1,\dots,b_d)=a_1b_1+\dots+a_db_d$ et muni de sa structure canonique de variété différentiable. La translation $\tau_m:n\to(n-m)$ identifie canoniquement $T(R^d)_m$ et R^d. Ceci permet de définir sur R^d une s.r. canonique en posant : $\langle x,y\rangle=\tau_m(x)\cdot\tau_m(y)$ quels que soient $x,y\in T(R^d)_m$.

(B) Soit W une v.r. et $f:V\to W$ une application indéfiniment différentiable d'une variété indéfiniment différentiable V dans W telle que, quel que soit $m\in V$, la différentielle $df|_m$ de f en m soit *injective*. Alors V hérite canoniquement de W une s.r. en posant $\|x\|=\|(df)(x)\|$ pour tout $x\in T(V)$. Deux cas particuliers sont importants :

(*a*) f est un revêtement; on dira alors que $f:V\to W$ est un revêtement riemannien; f est localement une isométrie.

(*b*) (V,f) est une sous-variété de la v.r. R^d de (A). Le théorème de Nash [22] assure que cette situation extrinsèque n'est pas, en fait, plus particulière que la situation générale de la définition d'une v.r. au § 1. Lorsque dim $V=2$ et $d=3$, on obtient la source historique de la notion de v.r. Lorsque $V=S_r^d=\{a\in R^{d+1}\mid \|a\|=r\}$ $(r>0)$, pour l'application identique $f:S_r^d\to R^{d+1}$, on obtient sur cette sphère une s.r. canonique, la seule que l'on considérera dans la suite, à l'exception de la fin du § 6.

(C) Soit G un groupe de Lie. Une s.r. sur G est dite invariante à gauche (resp. à droite) si toutes les translations à gauche λ_g (resp. à droite ϱ_g) sont

des isométries; une s.r. est dite biinvariante si elle est invariante et à droite et à gauche. On obtient sur G des s.r. invariantes à gauche en prenant sur $T(G)_e$ n'importe quelle structure euclidienne $a \cdot b$ et en posant $\langle x,y \rangle = (d\lambda_g^{-1})(x) \cdot (d\lambda_g^{-1})(y)$ quels que soient $x,y \in T(G)_g$. Si G est de plus compact on obtiendra sur G une s.r. biinvariante en prenant la moyenne, à droite et pour la mesure de Haar de G, d'une s.r. invariante à gauche.

Soit maintenant H un sous-groupe fermé du groupe de Lie compact G; une s.r. biinvariante sur G définit canoniquement une s.r. invariante par G sur l'espace homogène G/H (dite alors *naturelle*). On procède comme suit:soit $p: G \to G/H$ la projection canonique, $m \in G/H$ et g tel que $p(g) = m$. Soit $T(G)_g = V_g + W_g$ la décomposition orthogonale où V_g est le sous-espace de $T(G)_g$ constitué par l'espace tangent à la fibre gH. La restriction $dp: W_g \to T(G/H)_m$ est bijective et définit donc sur $T(G/H)_m$ une structure euclidienne. Comme la s.r. sur G est biinvariante, la structure euclidienne ainsi obtenue ne dépend pas du g choisi tel que $p(g) = m$ et l'on obtient ainsi sur G/H une s.r. canonique à partir de la s.r. biinvariante choisie sur G, et invariante par les opérations à gauche de G sur G/H.

Cas particuliers. Les espaces symétriques simplement connexes de rang un : $SO(d+1)/SO(d)$, $P^n(C) = SU(n+1)/U(n)$, $P^n(H) = Sp(n+1)/Sp(n) \times Sp(1)$, $P^2(\mathrm{Ca}) = F_4/\mathrm{Spin}$ [9]. Les s.r. ainsi obtenues sur l'un d'eux sont toutes proportionnelles; pour $SO(d+1)/SO(d)$ elles coïncident avec celles S_r^d de l'exemple $(B)(b)$. Pour $P^n(C)$, on obtient les métriques de Fubini-Study.

3. Notions classiques

Désormais V est une v.r. donnée. Si $\lambda: [0,1] \to V$ est une courbe de V, de vecteur tangent $\lambda'(t)$ en $\lambda(t)$, on définit sa longueur $L(\lambda) = \int_0^1 \|\lambda'(t)\| \cdot dt$. On pose $C(m,n) = \{\lambda \mid \lambda(0) = m \text{ et } \lambda(1) = n\}$. On a sur V une structure canonique d'espace métrique en posant : $d(m,n) = \inf_{\lambda \in C(m,n)} L(\lambda)$; la topologie correspondante coïncide avec la topologie initiale de la variété différentiable V. Pour étudier l'ensemble $SC(m,n) = \{\lambda \in C(m,n) \mid L(\lambda) = d(m,n)\}$ on est naturellement conduit à introduire une opération $D: \mathfrak{X}(V) \to \{\mathfrak{X}(V) \to \mathfrak{X}(V)\}$ qui, si elle est notée $X \to \{Y \to D_X Y\}$, satisfait les deux axiomes : $X(\langle Y,Z \rangle) = \langle D_X Y, Z \rangle + \langle Y, D_X Z \rangle$ et $D_X Y - D_Y X = [X, Y]$ quels que soient $X, Y, Z \in \mathfrak{X}(V)$. Il se trouve que D existe, est unique et de plus une loi de dérivation, c'est à dire que $D \in \mathrm{Hom}_{F^\infty(V)}(\mathfrak{X}(V); \mathrm{Hom}_R(\mathfrak{X}(V); \mathfrak{X}(V)))$ et $D_X(fY) = X(f) \cdot Y + f \cdot D_X Y$ quels que soient $X, Y \in \mathfrak{X}(V)$ et $f \in F^\infty(V)$. On dit que D est la loi de dérivation canonique de la v.r. V Le commutateur $R(X, Y) = D_X D_Y - D_Y D_X - D_{[X,Y]}$ se trouve être un tenseur de type (1,3) de V, appelé le tenseur de courbure de V. Il vérifie les identités classiques :

$$\langle R(x,y)z, t \rangle = -\langle R(x,y)t, z \rangle = \langle R(z,t)x, y \rangle, \tag{1}$$

$$R(x,y)z + R(y,z)x + R(z,x)y = 0 \tag{2}$$

et son tenseur dérivé DR vérifie :

$$(DR)(x,y;z) + (DR)(y,z;x) + (DR)((z,x;y) = 0. \tag{3}$$

On pourra aimer remarquer que, si l'on prend un système de coordonnées $\{\xi_i\}$ sur un ouvert U de V et si l'on pose $\partial/(\partial \xi i) = X_i \in \mathfrak{X}(U)$, la quantité $\langle R(X_i, X_j)X_k, X_h \rangle$ n'est autre que celle R_{ijkh} de la littérature classique.

Etant donnée $\lambda \in C(m,n)$, on définit le transport parallèle le long de λ comme l'application $\tau_\lambda : T(V)_m \to T(V)_n$ obtenue en exigeant que, $\{x(t)\}$ étant un champ de vecteurs le long de λ tel que $D_{\lambda'(t)}\, x(t) = 0$ quel que soit t, alors $\tau_\lambda(x(0)) = x(1)$. D'après le premier axiome de définition de D, il est clair que τ_λ est une isométrie d'espaces euclidiens. Une élément γ de $SC(m,n)$ devant nécessairement vérifier $D_{\gamma'(t)}\gamma'(t) = 0$ quel que soit t, une courbe γ telle que $D_{\gamma'(t)}\gamma'(t) = 0$ quel que soit t sera appelée une géodésique de V; à part les géodésiques triviales, on paramétrise dorénavant toutes les géodésiques en sorte que $\|\gamma'(t)\| = 1$ quel que soit t. Etant donné $x \in T(V)_m^*$ (i.e. $x \in T(V)_m$ et $x \neq 0$) il existe une géodésique et une seule γ_x telle que $\gamma_x(0) = p(x)$ et $\gamma_x'(0) = x/\|x\|$. Le théorème de Hopf-Rinow [20] assure que, si V est un espace topologique complet, alors $\gamma_x(t)$ existe quel que soit $x \in T(V)_m^*$, quel que soit $m \in V$ et quel que soit $t \in R$ et que, de plus, quels que soient $m, n \in V : SC(m,n) \neq \emptyset$ (évidemment $SC(m,n)$ peut avoir plus d'un élément). *Dorénavant toutes les* v.r. *considérées seront* COMPLÈTES.

4. La courbure sectionnelle

Pour découvrir si une v.r. V est isométrique ou non à R^d, pensant à $S^d \subset R^{d+1}$, il est naturel de procéder ainsi:soit $P \subset T(V)_m$ un sous-espace vectoriel de dimension deux de $T(V)_m$ (on désignera par $G(V)$ l'ensemble de tous ces P lorsque m parcourt V). Soit alors $C(P;\varepsilon)$ le petit cercle de V de direction P et de rayon ε, c'est-à-dire $C(P;\varepsilon) = \{\gamma_x(\varepsilon) \mid x \in T(V)_m^* \cap P\}$. Il s'agit d'étudier $l(P;\varepsilon) = L(C(P;\varepsilon))$. Il faut donc calculer le vecteur tangent à $C(P;\varepsilon)$ au point $\gamma_x(\varepsilon)$. Il est facile de voir que si $\{x,y\}$ est un repère orthonormé de P, ce vecteur tangent $Y(x,y;\varepsilon)$ est, après identifications convenables, la valeur $Z(\varepsilon)$ de la solution Z de l'équation différentielle dans $T(V)_m$:

$$\frac{d^2Z(t)}{dt^2} = \tau_t^{-1}(R(\gamma_x'(t), \tau_t(Z(t))\gamma_x'(t)) \tag{4}$$

telle que $Z(0) = 0$ et $dZ/dt|_{t=0} = y$ et où τ_t désigne le transport parallèle le long de γ_x de $\gamma_x(0) = m$ à $\gamma_x(t)$. On obtient donc pour $\|Y(x,y;\varepsilon)\|$ l'approximation :

$$\|Y(x,y;\varepsilon)\| = \varepsilon + \frac{\varepsilon^3}{6} \cdot \langle R(x,y)x, y\rangle + o(\varepsilon^3). \tag{5}$$

Il se trouve, grâce aux identités (1) et (2) que $\langle R(x,y)x,y\rangle$ ne dépend que de P et non du repère orthonormé choisi dans P. Puisque, lorsque $d=2$, $-\langle R(x,y)x,y\rangle$ n'est autre que la courbure totale de la surface V, il est naturel de définir la *courbure sectionnelle de V pour P* par $c(x,y) = c(P) = -\langle R(x,y)x,y\rangle$ pour tout repère orthonormé $\{x,y\}$ de P. On aura donc $c(P) = \lim_{\varepsilon \to 0} \frac{3}{\pi} \cdot (2\pi\varepsilon - l(P;\varepsilon))/\varepsilon^3$. On posera $c(V) = \{c(P) \mid P \in G(V)\}$ et on l'appellera l'*ensemble de courbure* de V. La connaissance de la fonction réelle $P \to c(P)$ sur $G(V)$ entraîne celle du tenseur de courbure lui-même.

Exemples. (A) — On a $c(R^d) = \{0\}$ (puisque $l(P;\varepsilon) = 2\pi\varepsilon$ ou bien parce que $R(X,Y) = 0$ quels que soient X, Y).

(*B*) (*a*) Si $f: V \to W$ est un revêtement riemannien, alors $c(V) = c(W)$ puisque f est localement une isométrie. (*b*) On a $c(S_r^d) = \{1/r^2\}$ (longueur d'un petit cercle). On sait aussi que, quel soit $\Delta < 0$, il existe sur R^d une s.r., notée $\mathfrak{a}^d(\Delta)$ et dite hyperbolique, telle que $c(\mathfrak{a}^d(\Delta)) = \{\Delta\}$. *On posera aussi, en* v.r. : $\mathfrak{a}^d(0) = R^d$ et, si $\Delta > 0 : \mathfrak{a}^d(\Delta) = S^d_{1/\sqrt{\Delta}}$.

(C) — On montre que, pour toute s.r. naturelle sur G/H, on a $c(G/H) \geqslant 0$. Plus précisément, si l'on identifie $T(G/H)_{p(e)}$ avec le W_e de la décomposition orthogonale $T(G)_e = V_e + W_e$ et $T(G)_e$ avec l'algèbre de Lie de G, on peut définir le crochet dans $T(G/H)_{p(e)}$. On montre alors que $c(x,y) = 0$ est équivalent à $[x,y] = 0$. D'autre part, pour les espaces riemanniens symétriques de rang un listés à la fin du § 2, autre que les sphères, on a : $c(V) = [\Delta/4, \Delta]$ $(\Delta > 0)$.

5. La courbure de Ricci

Il est aussi naturel de calculer le volume $v(m;\varepsilon)$ de la sphère de V de centre m et de rayon ε, définie par $S(m;\varepsilon) = \{\gamma_x(\varepsilon) \mid x \in T(V)^*_m\}$ (on sait que, pour ε assez petit, on a $S(m;\varepsilon) = \{n \in V \mid d(m,n) = \varepsilon\}$). Si $d\sigma$ est l'élément de volume de la sphère unité S_1^{d-1} de $T(V)_m$, on aura $v(m;\varepsilon) = \int_{S_1^{d-1}} v(x;\varepsilon) d\sigma$, où $v(x;\varepsilon) = \| Y(x,y_1;\varepsilon) \wedge \ldots \wedge Y(x,y_{d-1};\varepsilon) \|$ pour une base orthonormée de la forme $\{x, y_1, \ldots, y_{d-1}\}$ de $T(V)_m$. De (5) on obtient :

$$Y(x, y_1; \varepsilon) \wedge \ldots \wedge Y(x, y_{d-1}; \varepsilon)$$
$$= (y_1 \wedge \ldots \wedge y_{d-1}) \left(\varepsilon^{d-1} - \frac{\varepsilon^{d+1}}{6} \cdot \sum_{i=1}^{d-1} c(x, y_i) + o(\varepsilon^{d+1}) \right).$$

Pour un x tel que $\|x\| = 1$, la quantité $rc(x) = \sum_{i=1}^{d-1} c(x,y_i)$ n'est autre que la trace de la forme quadratique $\{y \to -\langle R(x,y)x, y\rangle\}$, trace *qui est appelée la courbure de Ricci de V en x*. On posera $rc(V) = \{rc(x) \mid x \in T(V)$ et $\|x\| = 1\}$. En général, pour un m fixé, $rc(x)$ dépend de x. Pour $d = 2$, on a $rc(V) = c(V)$; pour $d = 3$, la connaissance de $rc(V)$ entraîne celle de $c(V)$, par exemple $rc(V) = \{\Delta\}$ entraîne $c(V) = \{\Delta/2\}$. En général, s'il existe $f \in F^\infty(V)$ telle que $rc(x) = f(m)\|x\|^2$ quel que soit $x \in V_m$, on déduit de (3) que f est nécessairement constante et donc que $rc(V) = \{\Delta\}$.

Exemples. Si $c(V) = \{\Delta\}$ alors $rc(V) = \{(d-1)\Delta\}$; mais un exemple moins trivial est un espace homogène G/H du type de l'exemple (*C*) du § 2 tel que de plus la représentation adjointe ad H de H dans $T(G/H)_{p(e)}$ soit *irréductible* : en effet cette représentation laisse invariante sur $T(G/H)_{p(e)}$ deux formes quadratiques : $\| \ \|$ et rc. Puisque $\| \ \|$ est définie positive et ad H irréductible, ceci implique que rc et $\| \ \|$ sont proportionnelles.

6. La courbure scalaire

On peut maintenant achever le calcul de $v(m;\varepsilon)$. Si $\sigma_d = \int_{S_1^{d-1}} d\sigma$, on aura :

$$v(m;\varepsilon) = \sigma_d \cdot \varepsilon^{d-1} - \frac{\varepsilon^{d+1}}{6} \cdot \int_{S_1^{d-1}} rc(x) \cdot d\sigma + o(\varepsilon^{d+1}).$$

Puisque $rc(x)$ est une forme quadratique en $x \in T(V)_m$, si *sa trace est notée* $sc(m)$, on aura $\int_{S_1^{d-1}} rc(x) \cdot d\sigma = (1/d) \cdot sc(m)$. La fonction réelle $sc: V \to R$ est appelée la *courbure scalaire*. Quelle que soit la base orthonormée $\{x_i\}$ de $T(V)_m$, on aura $sc(m) = \sum_{i \neq j} c(x_i, x_j)$. Et

$$sc(m) = \lim_{\varepsilon \to 0} 6d \cdot \frac{\sigma_d \cdot \varepsilon^{d-1} - v(m;\, \varepsilon)}{\varepsilon^{d+1}}.$$

Exemple. Soit G le groupe de Lie simple, compact, simplement connexe, de dimension trois (donc G est difféomorphe à S^3). A l'aide des formules de Maurer-Cartan et des équations de structure, il est facile de construire sur G, quel que soit $\Delta \in R$, une structure riemannienne G_Δ invariante à gauche et telle que $sc(G_\Delta) = \{\Delta\}$.

7. Les problèmes considérés

Nous nous proposons maintenant de donner ci-dessous un exposé de résultats connus sur les relations entre la topologie de V et des hypothèses sur l'un des ensembles $c(V)$, $rc(V)$, $sc(V)$. Les résultats seront énoncés sans démonstration et nous nous restreindrons à trois types d'hypothèses :

dans le § 8 : la courbure est constante;

dans le § 9 : la courbure a un signe constant;

dans le § 10 : la courbure varie dans un certain domaine.

Une remarque préliminaire est: soit V une v.r. telle que $c(V)$, ou $rc(V)$, ou $sc(V)$ satisfasse une certaine condition. Si $\tilde{V} \to V$ est le revêtement riemannien universel de V, alors (d'après § 2, exemple (B), (a)) : $c(\tilde{V})$, ou $rc(\tilde{V})$, ou $sc(\tilde{V})$ satisfait la même condition. Le problème étudié se coupe donc en deux :

problème I : étudier une telle variété lorsqu'elle est simplement connexe;

problème II : étudier le groupe fondamental d'une telle variété. Nous observerons cette coupure dans les § 8 et 9. Rappelons qu'implicitement toutes les variétés considérées sont des VARIÉTÉS RIEMANNIENNES INDÉFINIMENT DIFFÉRENTIABLES, CONNEXES ET COMPLÈTES, de dimension supérieure ou égale à deux.

8. Les variétés à courbure constante

Problème I

THÉORÈME 1 [15]. *Soit V simplement connexe telle que $c(V) = \{\Delta\}$. Alors V est isométrique à $\mathfrak{a}(\Delta)$.*

Rappelons le fait classique : si $d = 2$, il existe sur V compacte une s.r. telle que $c(V) = \{\Delta\}$ (pour un certain Δ). La généralisation est le :

THÉORÈME 2 [30]. *Soit V une* v.r. *compacte* ($d \geqslant 3$). *Il existe sur V une fonction partout strictement positive et $\Delta \in R$ telle que la nouvelle* s.r. *$\overline{V}$ sur V définie par $\overline{\| \ \|} = f \| \ \|$ vérifie $sc(\overline{V}) = \{\Delta\}$.*

(La démonstration est un résultat fin d'équation aux dérivées partielles, elliptique et non linéaire.)

Le Théorème 2 montre que la courbure scalaire a peu de signification topologique. En outre l'exemple du § 6 montre que, pour $d=3$ (et peut être pour tout $d \geqslant 3$) même le signe de la courbure scalaire n'a pas d'implication topologique (tandis que, pour $d=2$, la formule de Gauss-Bonnet relie ce signe à celui de la caractéristique d'Euler-Poincaré de V).

Entre les deux théorèmes extrêmes 1 et 2, comment se comporte l'hypothèse $rc(V)=\{\Delta\}$? Le § 5 montre d'abord que cette question ne se pose que pour $d \geqslant 4$. Est-ce qu'une telle hypothèse est forte on non ? Peu de résultats sont connus. Remarquons d'abord que $rc(V)=\{\Delta\}$ a certainement des implications :

THÉORÈME 3. *Soit V compacte telle que $d=4$ et $rc\ (V)=\{\Delta\}$. Alors la caractéristique d'Euler-Poincaré $\chi(V)$ est positive ou nulle.*

(Ce résultat se déduit facilement de [11]).

Notons aussi qu'on ne connait pas d'exemples de v.r. compacte V telle que $rc(V)=\{\Delta\}$, autres que les espaces homogènes riemanniens G/H avec ad H irréductible. Les deux théorèmes suivants semblent montrer que V compacte et $rc(V)=\{\Delta\}$ est une très forte restriction :

THÉORÈME 4 [4, 6]. *Soit V compacte orientable telle que $d=4$ et $rc(V)=\{\Delta\}$. Si de plus : a) il existe $\Delta>0$ tel que $c(V)\subset]\Delta/4,\Delta]$ alors V est isométrique à $\mathfrak{a}(\Delta')$ (pour un certain Δ'); b) si $c(V)\geqslant 0$ et si V est kählérienne, alors V est isométrique a $P^2(C)$ ou a $S^2\times S^2$.*

THÉORÈME 5 [5]. *La structure riemannienne $\mathfrak{a}(\Delta)$ sur S^d est isolée parmi les* s.r. *dont la courbure de Ricci est égale à la constante $(d-1)\Delta$.* (Ces deux théorèmes reposent essentiellement sur la formule (6.8) de [19].)

Problème II

Ce qui précède montre que le Problème II n'a guère de sens, actuellement, que lorsque le revêtement universel riemannien de V est un $\mathfrak{a}(\Delta)$. Le Théorème 1 montre alors que le problème considéré est purement algébrique. Cependant, dans le cas $\Delta\leqslant 0$, on ne sait rien de plus que les Théorèmes 11, 12, 13 du § 9. Pour $\Delta=0$, on a le :

THÉORÈME 6 [8]. *Soit V compacte telle que $c(V)=\{0\}$. Il existe un revêtement riemannien W de V tel que W soit un tore.*

Pour $\Delta>0$, le Théorème 14 montre d'abord que seul le cas d'impair est à considérer. Lorsque $d=3$, le problème est complètement résolu par [25]. Pour $d\geqslant 5$, surtout en fait lorsque d est de la forme $4k+1$, le problème a été ramené dans [28] a un problème de pure arithmétique. Un résultat partiel mais complet est le :

THÉORÈME 7 [29]. *Soit V telle que $c(V)=\{\Delta\}$ $(\Delta>0)$ et homogène. Alors $\pi_1(V)$ est, soit cyclique, soit polyhédral binaire.*

(Le point de départ est de remarquer que l'homogénéité implique que les isométries de $\mathfrak{a}(\Delta)$ qui réalisent $\pi_1(V)$ sont des translations. On prouve alors le résultat algébrique conjecturé par Vincent.)

9. Les variétés dont la courbure a un signe constant

Problème I

On a déjà vu au § 8 qu'il n'a de sens que pour $c(V)$ ou $rc(V)$. On n'a de résultat en fait que pour $c(V)$ (voir cependant Théorème 17). C'est un problème naturel de rechercher les variétés topologiques qui admettent des s.r. telles que $c(V) \geqslant 0$, ou $c(V) \leqslant 0$. Du côté négatif ou nul, le problème est complètement résolu par le :

THÉORÈME 8 [14, 10]. *Soit V simplement connexe telle que $c(V) \leqslant 0$. Alors V difféomorphique à R^d.*

(La démonstration repose essentiellement sur le théorème de Hopf-Rinow et le fait que, grâce à (4), l'application exponentielle $\exp_m : T(V)_m \to V$ (pour un m quelconque) a une différentielle partout bijective, et est alors un revêtement).

Du côté positif ou nul, les exemples du § 4 (C) montrent que le problème n'est pas aussi simple (excepté en dimension deux : voir § 8). Même le cas $c(V) > 0$ est difficile (exemples (C) du § 4). En fait on ne connait actuellement aucune implication topologique de la condition $c(V) > 0$. Quelques résultats partiels :

THÉORÈME 9 [13]. *Soit V kählérienne compacte telle que $c(V) > 0$. Alors deux sous-variétés analytiques complexes compactes, de dimension complémentaires, de V se rencontrent nécessairement. Si de plus $d = 4$, ceci implique que V est homéomorphe à $P^2(C)$.*

(La démonstration de la première partie utilise un argument de minimum et la variation seconde; la deuxième partie est due à Andreotti.)

THÉORÈME 10 [3]. *Soit G/H un espace homogène riemannien naturel simplement connexe et tel que $c(G/H) > 0$. A part deux exceptions (l'une pour $d = 7$, l'autre pour $d = 13$), la structure variété sous-jacente de G/H est homéomorphe a l'une des : S^d, $P^{(d/2)}(C)$, $P^{(d/4)})(H)$, $P^2(Ca)$.*

(La démonstration consiste à utiliser ce qui a été dit dans l'exemple (C) du § 4 pour se ramener à un problème d'algèbres de Lie.)

Problème II

THÉORÈME 11 [10]. *Soit V telle que $c(V) \leqslant 0$. Alors tout sous-groupe de $\pi_1(V)$ est infini.*

(La méthode consiste à montrer que $c(V) \leqslant 0$ entraîne qu'un compact de $\tilde{V}$ a un barycentre unique; on utilise alors comme compact l'orbite d'un sous-groupe fini de $\pi_1(V)$.)

THÉORÈME 12 [23]. *Soit V compacte telle que $c(V) < 0$. Alors $\pi_1(V)$ ne peut pas être abélien et tous ses sous-groupes abéliens sont cycliques.*

(La démonstration utilise l'existence d'une géodésique fermée dans une classe d'homotopie libre non nulle de V et le fait que $c(V) < 0$ entraîne que tout quadrilatère de V a des angles dont la somme est strictement inférieure à 2π).

THÉORÈME 13 [18]. *Soit V une variété riemannienne homogène telle que $c(V)<0$. Alors V est simplement connexe.*

(Utilise une géodésique fermée comme ci-dessus et un champ de Jacobi-Killing le long d'icelle : contradiction si $c(V)<0$.)

On trouvera dans [12] des résultats fins sur les V telles que $d=2$, $c(V)>0$, V non compacte. On a un résultat définitif :

THÉORÈME 14 [26]. *Soit V compacte, orientable, de dimension paire et telle que $c(V)>0$. Alors V est simplement connexe.*

(On applique la variation seconde le long d'une plus courte géodésique fermée).

Et des résultats partiels :

THÉORÈME 15 [21]. *Soit V compacte telle que $rc(V)>0$. Alors $\pi_1(V)$ est fini.*

(La variation seconde montre que $rc(V)\geqslant\delta>0$ implique que V est de diamètre fini donc compacte. On applique ceci à $\tilde{V}$.)

THÉORÈME 16 [11]. *Soit V compacte telle que $d=4$. Alors, si $c(V)\geqslant 0$ ou si $c(V)\leqslant 0$, on a $\chi(V)\geqslant 0$; et $\chi(V)=0$ implique $c(V)=\{0\}$.*

(S'obtient en choisissant un repère orthonormé tel que l'expression figurant sous le signe somme dans la formule généralisée de Gauss-Bonnet [1] ait une forme simple faisant intervenir la courbure sectionnelle.)

THÉORÈME 17 [9]. *Soit V kählérienne compacte telle que $rc(V)>0$. Alors il n'existe pas sur V de formes extérieures holomorphes fermées non triviales.*

(La méthode est celle des formes harmoniques : voir formule (3.4) de [19].)

THÉORÈME 18 [17]. *Soit V kählérienne compacte telle que $rc(V)>0$. Alors V est simplement connexe.*

(On utilise simultanément : le Théoreme 15, le fait que V est algébrique d'après le théorème de plongement de Kodaira et les genres arithmétiques de V et $\tilde{V}$).

10. Les variétés k-pincées

On introduit une notion plus forte que $c(V)>0$. *Une* v.r. *est dite k-pincée s'il existe $\Delta>0$ et k tels que $c(V)\subset[k\Delta,\Delta]$.*

(Seul le Problème I a été considéré. Pour $k=\frac{1}{4}$ les résultats sont satisfaisant (seule la question du difféomorphisme restant ouverte)).

THÉORÈME 19 [24, 16]. *Soit V simplement connexe et k-pincée. Si $k>\frac{1}{4}$, alors V est homéomorphe à S^d.*

(La démonstration repose essentiellement sur deux résultats : celui de [16] sur le cut-locus et celui de [27] qui est essentiellement une inégalité trigonométrique).

THÉORÈME 20 [2]. *Soit V simplement connexe, non homéomorphe à S^d et $(\frac{1}{4})$-pincée. Alors V est isométrique a l'un des : $P^{(d/2)}(C)$, $P^{(d/4)}(H)$, $P^2(Ca)$.*

(Raffinant la démonstration du Théorème 19, on montre que la courbure sectionnelle à travers une géodésique quelconque est invariante par transport parallèle le long d'elle; d'après [10], ceci implique que V est une v.r. symétrique).

Bien que son existence théorique soit probable, on ne connait pas actuellement un ε $(0<\varepsilon<(\frac{1}{4}))$ tel que l'hypothèse « V est simplement connexe et k-pincée avec $k>(\frac{1}{4})-\varepsilon$ » entraîne « V est homéomorphe à l'une des : S^d, $P^{(d/2)}(C)$, $P^{(d/4)}(H)$, $P^2(Ca)$ » (on remarquera les deux exceptions du Théorème 10). On a un résultat partiel :

THÉORÈME 21 [2]. *Soit V compacte telle que $d=2n+1$ $(n\geqslant 2)$ et k-pincée. Alors si $k=2(n-1)/(8n-5)$, le deuxième nombre de Betti réel de V est nul.*

(La méthode utilise la formule (3.4) de [19]).

En particulier si $d=5$ et $k=2/11$, alors V aura la cohomologie réelle de S^5. Un seul résultat (très faible comme le montrent la substitution $k=\frac{1}{4}$ et le Théorème 20) est connu reliant l'hypothèse k-pincée et la topologie de V:

THÉORÈME 22 [7]. *Soit V compacte, k-pincée avec $k>0$ et telle que $d=2n$. Alors : $|\chi(V)|\leqslant 2^{-n}\cdot(2n)!\cdot k^{-n}$.*

(La méthode utilise la formule généralisée de Gauss-Bonnet [1] et les majorations pour le tenseur de courbure de [2].)

BIBLIOGRAPHIE

[1]. ALLENDOERFER, C. & WEIL, A., The Gauss- Bonnet theorem for Riemannian polyhedra. *Trans. Amer. Math. Soc.*, 53 (1943), 101–129.

[2]. BERGER, M., Sur quelques variétés riemanniennes suffisamment pincées. *Bull. Soc. Math. France*, 88 (1960), 57–71.

[3]. —— Les variétés riemanniennes homogènes normales simplement connexes à courbure strictement positive. *Ann. Scuola Norm. Sup. Pisa*, XV 1961, 179–246).

[4]. —— Sur quelques variétés d'Einstein compactes. *Ann. di Mat.* 53 (1961), 89–96.

[5]. —— Les sphères parmi les variétés d'Einstein. *C. R. Acad. Sci. Paris*, 254 (1962), 1564–1566.

[6]. —— Les variétés kählériennes d'Einstein de dimension quatre à courbure positive (A paraître dans *Tensor*.)

[7]. —— On the characteristic of positively-pinched Riemannian manifolds. (A paraître 1962 dans *Proc. Nat. Acad. Sci. U.S.A.*)

[8]. BIEBERBACH, Über die Bewegungsgruppen der Euklidischen Raüme. *Math. Ann.* 72 (1912), 400–412.

[9]. BOCHNER, S., Tensorfields and Ricci curvature in Hermitian metric. *Proc. Nat. Acad. Sci. U.S.A.*, 37 (1951), 704–706.

[10]. CARTAN, E., *Leçons sur la géométrie des espaces de Riemann*. Paris, 1951.

[11]. CHERN, S. S., On curvature and characteristic classes of a Riemannian manifold *Abh. Math. Sem. Hamburg*, 20 (1955), 117–126.

[12]. COHN-VOSSEN, Totalkrümmung und Geodätische Linien auf einfach zusammenhängenden, offenenvollständigen Flächenstücken. *Mat. Sb. N.S.*, 1 (1936). 139–164.

[13]. FRANKEL, T., Manifolds with positive curvature. *Pacific J. Math.*, 11 (1961), 165–174.

[14]. HADAMARD, J., Les surfaces à courbures opposées et leur lignes géodésiques. *J. Math. Pures Appl.*, 4 (1898), 27–73.

[15]. HOPF, H., Zum Clifford-Kleinschen Raumproblem. *Math. Ann.*, 95 (1926), 313–339.

[16]. KLINGENBERG, W., Über Riemannsche Mannifaltigkeiten mit nach oben beschränkter Krümmung. (A paraître dans *Ann. di Mat.* 1962.)

[17]. KOBAYASHI, S., On compact Kähler manifolds with positive definite Ricci tensor. *Ann. Math.*, 74 (1961), 570–574.

[18]. —— Homogeneous Riemannian manifolds of negative curvature. *Bull. Amer. Math. Soc.*, (1962), 338–339.

[19]. LICHNEROWICZ, A., *Géométrie des Groupes de Transformations*. Paris, 1958).

[20]. MYERS, S., Riemannian manifolds in the large. *Duke Math. J.*, 1 (1935), 39–45.

[21]. —— Riemannian manifolds with positive mean curvature. *Duke Math. J.*, 8 (1941), 401–404.

[22]. NASH, J., The imbedding problem for Riemannian manifolds. *Ann. Math.*, 63 (1956), 20–63.

[23]. PREISSMANN, A., Quelques propriétés globales des espaces de Riemann. *Comment. Math. Helv.*, 15 (1942), 175–216.

[24]. RAUCH, H. E., *Geodesics and Curvature in Differential Geometry in the Large*. New York, Yeshiva University, 1959.

[25]. SEIFERT, H. & THRELLFALL, W., Topologische Untersuchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes. *Math. Ann.*, 107 (1932), 543–586.

[26]. SYNGE, J. L., On the connectivity of spaces with positive curvature. *Quart. J. Math. Oxford*, (1936), 316–320.

[27]. TOROHOGOV, V. A. Римановы пространства кривизны ограниченной снизу. *успехи мат. наук*, 85 (1959), 87–130.

[28]. VINCENT, G., Les groupes linéaires finis sans points fixes. *Comment. Math. Helv.*, 20 (1947), 117–171.

[29]. WOLF, J. A., Vincent's conjecture on Clifford translations of the sphere. *Comment. Math. Helv.*, 136 (1961), 33–41.

[30]. YAMABE, H., On deformation of Riemannian structures on compact manifolds. *Osaka Math. J.*, 12 (1960), 21–37.

EMBEDDING SURFACES IN 3-MANIFOLDS

By R. H. BING

Is a 2-sphere in E^3 tame if it can be pierced at each point by a straight line interval?(1) Is it tame if it can be pierced along each arc in it by a tame disk?

Is a 2-sphere in E^3 tame if it is homogeneous under a space homeomorphism?

Is a 2-sphere in E^3 tame if each Sierpinski curve in it can be pushed to either side? Is it tame if it can be approximated from either side by a singular 2-sphere?

Does one obtain E^3 if the interior of a wild 2-sphere is replaced by the interior of a tame one?

What restrictions imply that a surface is tamely embedded in a 3-manifold?

These are samples of some of the interesting unsolved problems about the embedding of surfaces in 3-manifolds.

In addition to raising questions about the embedding of surfaces in 3-manifolds, recent discoveries were discussed. A surface is known to be tame if its complement is 1-ULC. It is tame if it can be pushed to either side with a map. It is tame if it can be homeomorphically approximated from either side. The fact that surfaces can be approximated by polyheral surfaces has been useful in proving some of these results and may be the means for learning more about the embedding of surfaces in 3-manifolds.

One advantage of working with polyhedral surfaces instead of arbitrary ones is that they intersect in decent sets. However, it has been shown recently that if in E^3, L is a straight line, S is a 2-sphere, and $\varepsilon > 0$, then there is a homeomorphism h of E^3 onto itself such that h does not move any point more than ε, h is fixed outside an ε-neighborhood of L, and $S \cdot h(L)$ does not have infinitely many points. A simple proof of this result would improve the proof of the approximation theorems. Suppose S_1, S_2 are two arbitrary surfaces in E^3. One might wonder if it is possible to move S_1 slightly with a space homeomorphism so that the resulting 2-sphere intersects S_2 "nicely".

Hand in hand with proofs go counterexamples. Examples have recently been given of wild spheres all of whose arcs are tame. It has been shown however that all wild surfaces contain tame arcs and can be pierced by tame arcs. The set of points that belong to tame arcs in the surfaces is precisely the set of points of the surface at which it can be pierced by tame arcs.

Does a simple closed curve bound a disk in a 3-manifold if the curve can be shrunk to a point in its own complement? This is an unsolved version of Dehn's lemma.

If a singular plane P in E^3 links two points in E^3, does each ε-neighbor-

(1) Following the lecture, M. K. Fort, Jr. answered the first question raised in this lecture in the negative by giving an example of a wild 2-sphere (which might be called a "wild porcupine") which could be pierced at each point by a straight line interval. The example is described in a paper submitted to the *Proc. Amer. Math. Soc.*

hood of P contain a non-singular plane that links the two points? This version of the sphere theorem remains unanswered partly because in Papakyriakopoklos' proof of Dehn's lemma, the splitting and resewing along double lines introduced long feelers. Suppose D is a disk on a polyhedral 2-sphere S and $\varepsilon > 0$. Is there a positive number $\delta > 0$ such that if g is a piecewise linear map of S into E^3 that is fixed on $S - D$ and moves no point more than δ, then each neighborhood of $g(S)$ contains a 2-sphere whose homeomorphic distance from S is less than ε?

A contemplation of E^3 and a study of recent discoveries about 3-manifolds reveals that although we have learned much about 3-space in recent years, there is much yet to learn.

PARTITIONS OF UNITY IN THE THEORY OF FIBRATIONS

By ALBRECHT DOLD

Partitions of unity have been used on many occasions to deduce global results on spaces or maps from local properties. Below we state a section extension theorem of this type and indicate how it implies several of the main existence theorems for fibrations. This permits us to define a very smooth theory of fibre bundles. Another application is to fibre homotopy equivalence.

Definitions. Let B be a topological space. A *halo* around $A \subset B$ is a subset V of B such that there exists a continuous function $\tau: B \to [0,1]$ with $\tau | A = 1$, $\tau | CV = 0$. For example, every $V \subset B$ is a halo around $A = \phi$; if B is normal and A closed then every neighborhood of A is a halo.

A (not necessarily open) covering $\{V_\lambda\}_{\lambda \in \Lambda}$ of B is called *numerable* (or normal) if there exists a locally finite partition of unity $\{\pi_\gamma: B \to [0,1]_{\gamma \in \Gamma}$ such that every set $\pi_\gamma^{-1}(0, 1]$ is contained in some V_λ. For example, paracompact resp. normal spaces are characterized by the fact that every open covering resp. every locally finite open covering is numerable.

Definition. A continuous map $p: E \to B$ has the *section extension property* (SEP) if the following holds. For every $A \subset C$ and every section s of p over A which admits an extension to a halo V around A there exists an extension S over B, i.e. a section $S: B \to E$ with $S | A = s$.

SECTION EXTENSION THEOREM. *If there exists a numerable covering $\{V_\lambda\}_{\lambda \in \Lambda}$ of B such that $p_\lambda = p | p^{-1}(V_\lambda): p^{-1}(V_\lambda) \to V_\lambda$ has the SEP for every $\lambda \in \Lambda$ then p has the SEP.*

(Compare Godement, *Théorie des Faisceaux*, II, 3.)

The usual results on sections in fibrations are then contained in the following

COROLLARY. *Let $p: E \to B$ be a map and $\{V_\lambda\}_{\lambda \in \Lambda}$ a numerable covering of B such that $p_\lambda: p^{-1}(V_\lambda) \to V_\lambda$ is fibre homotopically trivial for every λ, say $p^{-1}(V_\lambda) \sim V_\lambda \times Y_\lambda$. If*

(a) Y_λ is contractible, or

(b) Y_λ is $(n-1)$-connected and V_λ a CW complex of dimension $\leqslant n$ for every $\lambda \in \Lambda$ then p has the SEP.

Definition. A map $p: E \to B$ has the *covering homotopy property* ($=CHP$) if for every $\bar{H}: X \times [0, 1] \to B$ and $h: X \to E$ with $ph(x) = \bar{H}(x, 0)$ there exists $H: X \times [0, 1] \to E$ with $pH = \bar{H}$ and $H(x, 0) = h(x)$.

If the existence of H is only required for homotopics $\bar{H}$ which satisfy $\bar{H}(x,t) = \bar{H}(x,0)$ for $t \leqslant \frac{1}{2}$ then p is said to have the *weak covering homotopy property* (WCHP).

The WCHP is useful because it is invariant under fibre homotopy equivalence whereas the CHP is not.

COVERING HOMOTOPY THEOREM. *Let $p:E\to B$ be a map and $\{V_\lambda\}_{\lambda\in\Lambda}$ a numerable covering of B such that $p_\lambda:p^{-1}(V_\lambda)\to V_\lambda$ has the CHP (the WCHP) for every $\lambda\in\Lambda$. Then p has the* CHP (*the* WCHP).

The CHP case is due to Hurewicz and Huebsch, the WCHP has also been found by Weinzweig (unpublished). For the proof one introduces an auxiliary space

$$\tilde{E}=\{(x,w)\in X\times E^{[0,1]}\,|\,w(0)=h(x) \text{ and } pw(t)=\bar{H}(x,t)\}$$

and the map

$$\tilde{p}:\tilde{E}\to X \quad (\tilde{p}(x,w)=x)$$

A section for $\tilde{p}$ is then the same as a covering homotopy H for $\bar{H}$ with initial position h.

The notion of *fibre homotopy equivalence* makes sense for arbitrary continuous maps $p:E\to B$, $p':E'\to B$. Using an auxiliary construction as above the SE theorem here gives:

THEOREM. *Let $f:E\to E'$ be a continuous map with $p'f=p$. Assume there is a numerable covering $\{V_\lambda\}_{\lambda\in\Lambda}$ of B such that $f_\lambda:p^{-1}(V_\lambda)\to p'^{-1}(V_\lambda)$ is a fibre homotopy equivalence for every $\lambda\in\Lambda$. Then f is a fibre homotopy equivalence.*

The role of the WCHP for fibre homotopy equivalence is made clear by the following results, parts of which have also been found by Fadell, Fuchs and Weinzweig.

THEOREM. *Assume B admits a numerable covering $\{V_\lambda\}_{\lambda\in\Lambda}$ such that the inclusion map $V_\lambda\to B$ is nullhomotopic for every $\lambda\in\Lambda$. Then $p:E\to B$ has the* WCHP *if and only if $p_\lambda:p^{-1}(V_\lambda)\to V_\lambda$ is fibre homotopically trivial for every λ.*

THEOREM. *Let B, $\{V_\lambda\}_{\lambda\in\Lambda}$ be as above, assume $p:E\to B$, $p':E'\to B$ have the* WCHP *and $f:E\to E'$ satisfies $p'f=p$. Then the following are equivalent.*

(*a*) *f is a fibre homotopy equivalence,*
(*b*) *f is a (ordinary) homotopy equivalence*
(*c*) *the restriction $f_b:p^{-1}(b)\to p'^{-1}(b)$ of f is a homotopy equivalence for at least one point b in every component of B.*

The applications to bundle theory are based on the notion of numerable bundle. A fibre bundle ξ is called *numerable* if there exists a numerable covering $\{V_\lambda\}_{\lambda\in\Lambda}$ of $B=B_\xi$ (the base of ξ) such that $\xi\,|\,V_\lambda$ (the part of ξ over V_λ) is trivial for every $\lambda\in\Lambda$. The class C of numerable bundles seems to have the right size for a satisfactory bundle theory, namely:

I. *It is large enough:* If B_ξ is paracompact then ξ is always numerable. Milnor's universal bundle is numerable. If ξ is numerable then all bundles which are induced from or associated with ξ are numerable. If H is a subgroup of $G=G_\xi$ (the structure group of ξ) then every H-bundle which is obtained from ξ by reduction of the structure group is numerable, provided the coset-bundle $G\to G/H$ is numerable.

II. *It is small enough:* The covering homotopy theorem, the section theorem for bundles with contractible fibre, and the classification theorem hold

in C without restriction. In order to formulate the latter we denote by $k_G(X)$ the set of equivalence classes of numerable (principal) G-bundles with base X. Such a bundle η is called *universal* if

$$T_\eta : \pi[Y, B_\eta] \to k_G(Y),$$

$$T_\eta[f] = f^*(\eta)$$

is bijective for all topological spaces Y, where π [ ] denotes homotopy classes of maps $f: Y \to B\eta$. The classification theorem then asserts that η is universal if and only if E_η, its total space, is contractible. The existence of universal bundles is proved by Milnor's construction.

A continuous homomorphism $\gamma: G \to H$ of topological groups induces a natural transformation $k_\gamma: k_G \to k_H$ (the "associated bundle") and one proves that k_γ is an equivalence if and only if γ is a (ordinary) homotopy equivalence.

SOME CONNECTIONS BETWEEN COBORDISM AND TRANSFORMATION GROUPS

By E. E. FLOYD

Consider as a general theme differentiable actions of a compact Lie group G on a closed differentiable manifold V^n. It being difficult to handle complicated structure on both G and V^n simultaneously, here G will always be of very easy structure. P. E. Conner and I have studied in some detail the connections between the cobordism class of V^n and properties of transformation groups acting on V^n. Here I wish to discuss an aspect of our work that interests me particularly: given a compact Lie group G, which cobordism classes admit representatives upon which G acts differentiably and without stationary points (x is *stationary* if $gx=x$ for all $g \in G$)? Full details will be forthcoming in a joint memoir of Conner and myself; I hope that the present remarks give some insight into the problems we encounter and the methods we use.

Make the agreement here that all actions are differentiable of class C^∞, and that all manifolds V^n are compact differentiable; V^n is closed if it is without boundary.

1. Actions of $(Z_2)^k$

Consider first actions of Z_2: that is, consider differentiable involutions $T: V^n \to V^n$. If T is without fixed points, what is the cobordism class of V^n? Recall that the closed n-manifold V^n is said to *cobord* (mod 2) if there exists a compact manifold B^{n+1} with $\dot{B}^{n+1} = V^n$. It is our contention that if a differentiable involution T acts on V^n without fixed points, then V^n cobords. The proof is easy. One considers the orbit map $\nu: V^n \to V^n/T$ and the mapping cylinder B^{n+1} of ν. A little thought reveals that B^{n+1} is a compact manifold with $\dot{B}^{n+1} = V^n$. Hence our problem is solved for $G = Z_2$.

For any compact Lie group G, consider all pairs (G, V^n) consisting of a closed manifold V^n and a free action of G on V^n. Two such, (G, V_1^n) and (G, V_2^n), are cobordant if there exists a free action (G, B^{n+1}) on an $(n+1)$-manifold with

$$(G, \dot{B}^{n+1}) = (G, V_1^n) \cup (G, V_2^n).$$

There results an abelian group $\mathfrak{n}_n(G)$ of cobordism classes; denote the class containing (G, V^n) by $[G, V^n]_2$. There is the Thom algebra $\mathfrak{n} = \Sigma \mathfrak{n}_n$ of ordinary unoriented cobordism classes. Now $\mathfrak{n}_*(G) = \Sigma \mathfrak{n}_n(G)$ is an $\mathfrak{n}$-module, where

$$[G, V^m]_2 \cdot [W^n]_2 = [G, V^m \times W^n]_2$$

with $(G, V^m \times W^n)$ given by $g(x,y) = (gx, y)$. It turns out, using the methods of Thom [5], that $\mathfrak{n}_*(G)$ is always a free $\mathfrak{n}$-module; in fact,

$$n_*(G) \simeq H_*(B_G; Z_2) \otimes n,$$

where B_G is a classifying space for G.

In particular, $n_*(Z_2)$ is a free n-module with generators $(Z_2, S^n]$, $n = 0, 1, \ldots$, where Z_2 acts on S^n via the antipodal map. Moreover,

$$n_*((Z_2)^k) \simeq n_*(Z_2) \otimes_n \ldots \otimes_n n_*(Z_2).$$

With these preliminaries, consider now actions of $(Z_2)^k$. Equivalently, consider differentiable involutions $T_i: V^n \to V^n$, $i = 1, \ldots, k$, with $T_i T_j = T_j T_i$. We outline here the following theorem.

((1.1) *If $(Z_2)^k$ acts on the closed manifold V^n without stationary points, then V^n cobords mod* 2.

Proof. The basic technique of the proof is to define inductively a sequence

$$V^n = V_0^n, V_1^n, \ldots, V_k^n$$

of closed manifolds, each carrying an action of $(Z_2)^k$. We call the sequence the *free resolution* of the action $((Z_2)^k, V^n)$.

Suppose the action $((Z_2)^k, V_j^n)$ has been defined, $j < k$. Fix a Riemannian metric on V_j^n, invariant under the action. Consider $R \times V_j^n$, R the real numbers, with the action given by

$$T_i(t, x) = (t, T_i(x)), i \neq j+1, T_{j+1}(t, x) = (-t, T_{j+1}(x)).$$

The fixed point set $F(T_{j+1}, R \times V_j^n)$ is then $O \times F(T_{j+1}, V_j^n)$. Let N denote a tubular neighborhood of $F(T_{j+1}, R \times V_j^n)$ of small radius ε; then N is invariant under the action. We let V_{j+1}^n be the boundary $\dot{N}$.(1)

Note that T_{j+1} operates on V_{j+1}^n without fixed points. Closer inspection reveals in fact that the group $(Z_2)^{j+1}$ generated by $T_i, \ldots, T_{j+1}$ acts freely on V_{j+1}^n. Moreover, V_{j+1}^n is the bundle space of sphere bundle $V_{j+1}^n \to F(T_{j+1}, V_j^n)$, with the bundle map equivariant.

Given the action $((Z_2)^k V^n)$, we thus obtain the free action $((Z_2)^k V_k^n)$. The heart of the proof lies in the relationship between these two actions. Consider $\gamma = [(Z_2)^k (Z_2)^k]_2 \in n_0(Z_2)^k)$, where $(Z_2)^k$ acts on itself by left multiplication. We prove by an elementary geometric argument that $[(Z_2)^k V_k^n]_2 = \gamma \cdot [V^n]_2$ in $n_*((Z_2)^k)$.

Suppose now that $(Z_2)^k$ acts on V^n without stationary points. We argue that $V_k^n = \emptyset$. There is the composite map

$$V_k^n \to F(T_k, V_{k-1}^n) \subset V_{k-1}^n \to F(T_{k-1}, V_{k-2}^n) \subset V_{k-2}^n \ldots \subset V_0^n = V^n,$$

whose image is seen to be the set of stationary points. Hence $V_k^n = \emptyset$ and $\gamma \cdot [V^n]_2 = 0$. But γ is a non-zero element of the free n-module $n_*((Z_2)^k)$, hence $[V^n]_2 = 0$.

(1) (Added in proofs): In the proof of (1.1), $F(T_{j+1}, V_j^n)$ must be replaced by the disjoint union $\bigcup F(T^{i(1)} \ldots T^{i(j)} T^{j+1}, V_j^n)$. The outline of proof is otherwise correct. A similar change must be made in section 2.

2. Actions of $(Z_p)^k$

Consider hereafter the oriented case. Namely all manifolds are oriented, and all group actions preserve the orientation. We use now the oriented cobordism ring $\Omega = \Sigma\Omega_n$ of Thom [5]. There is the cobordism module $\Omega_*(G) = \Sigma\Omega_n(G)$, defined as was $n_*(G)$ except that the orientation conventions are imposed.

Denote by $SF_n(G) \subset \Omega_n$ the set of all cobordism classes which admit representatives V^n upon which G acts without stationary points; SF here abbreviates "stationary point free". $SF_n(G)$ is seen to be a subgroup of Ω_n. Moreover $SF(G) = \Sigma\, SF_n(G)$ is an ideal in Ω. For if G acts without stationary points on V^m then G acts without stationary points on $V^m \times W^n$ by $g(x,y) = (gx, y)$.

It follows from section 1 that $SF((Z_2)^k) = 2\Omega$. A corresponding result holds for Z_p, p prime.

(2.1) *$SF(Z_p)$, p prime, is the ideal $p\Omega$ of Ω.*

Proof. If Z_p acts without stationary points on V^n, then Z_p acts freely on V^n. Assume now the consequence of the work of Milnor [4] and Wall [6] that the Pontryagin numbers and Whitney numbers of V^n determine $[V^n]$. It can now be proved that $[V^n] = p[V^n/Z_p]$. This follows readily from a comparison of Pontryagin numbers and Whitney numbers. The remark follows from this fact.

We might remark at this stage that we have considerable information about $\Omega_*(Z_p)$, p an odd prime. This Ω-module is generated by the elements $\alpha_{2k-1} = [Z_p, S^{2k-1}]$, where the generator T of Z_p acts on S^{2k-1} by $T(z_1, \ldots, z_k) = (\varrho z_1, \ldots, \varrho z_k)$ with $\varrho = \exp 2\pi i/p$, and by $\alpha_0 = [Z_p, Z_p]$. Let F be the free Ω-module generated by $\alpha_0, \alpha_1, \alpha_3, \ldots$ and let R be the submodule generated by the elements

$$\beta_{2k-1} = p\,\alpha_{2k-1} + \alpha_{2k-5} \cdot [W^4] + \alpha_{2k-9}[W^8] + \ldots.$$

For appropriate choice of $W^4, W^8, \ldots, F/R \simeq \Omega_*(Z_p)$. Moreover, $[W^4], \ldots, [W^{2p-6}]$ are all divisible by p and $[W^{2p-2}] = b[P_{p-1}(C)]$ in $\Omega/p\Omega$ where $b \neq 0 \mod p$.

Now we continue (2.1) by considering $SF((Z_p)^k)$. Here we have been motivated strongly by the outstanding work of Borel [1,2] on actions of $(Z_p)^k$. In particular we start with some examples of Borel. Note first that $Z_p \times Z_p$ acts without stationary points on complex projective space $P_{p-1}(C)$, with generators given by

$$T_1[z_i, \ldots, z_p] = [z_2, \ldots, z_p, z_1],\ T_2[z_1, \ldots, z_p] = [z_1, \varrho z_2, \ldots, \varrho^{p-1} z_p]$$

where $\varrho = \exp(2\pi i/p)$. Since $[P_{p-1}(C)]$ is a generator of Ω, this shows that $SF(Z_p \times Z_p) \neq p\Omega$. We shall see later that, in a certain sense, this is the only example needed for $Z_p \times Z_p$. Borel has gone on to show that $Z_3 \times Z_3 \times Z_3$ acts without stationary points on the Cayley plane, but that $Z_3 \times Z_3$ cannot.

We pass now to our results. Suppose $(Z_p)^k$ is generated by $T_1, \ldots, T_k$ where $T_j^p = 1$ and $T_i T_j = T_j T_i$.

Consider an action $((Z_p)^k, V^n)$ p an odd prime. We generalize the constructions of § 1 to obtain a *free resolution*

$$V^n, V^{n+1}, \ldots, V^{n+k}$$

of actions $((Z_p)^k, V^{n+j})$. Consider $C \times V^{n+j}$, C the complex numbers, with the action given by

$$T_i(z,x) = (z, T_i x), i \neq j+1, T_{j+1}(z,x) = (\varrho z, T_{j+1}(x)) \quad \text{with}$$

$\varrho = \exp\ (2\pi i/p)$. The fixed point set $F(T_{j+1}, C \times V^{n+j})$ is $O \times F(T_{j+1}, V^{n+j})$. Let N denote a tubular neighborhood of $F(T_{j+1}, C \times V^{n+j})$ in $C \times V^{n+j}$ of small radius ε, and let $V^{n+j+1} = \dot{N}$. As before, the group $(Z_p)^{j+1}$ generated by $T_1, \ldots, T_{j+1}$ acts freely on V^{n+j+1}. Also as before, V^{n+j+1} is the bundle space of a sphere bundle $V^{n+j+1} \to F(T_{j+1}, V^{n+j})$, with the bundle map equivariant.

The following theorem uses the free action $((Z_p)^k, T^k)$ on the torus $T^k = S^1 \times \ldots \times S^1$ given by $T_j(z_1, \ldots, z_j, \ldots, z_k) = (z_1, \ldots, \varrho z_j, \ldots, z_k)$. Its proof is too long to sketch.

(2.2) *Consider an action* $((Z_p)^k, V^n)$, *p an odd prime, and consider a free resolution* $V^n, \ldots, V^{n+k}$. *We have*

$$[(Z_p)^k, V^{n+k}] = [(Z_p)^k, T^k]\,[V^n]$$

in $\Omega_{n+k}((Z_p)^k)$.

Suppose in particular that $(Z_p)^k$ acts on V^n without stationary points. As in section 1, then $V^{n+k} = \emptyset$. Hence $[(Z_p)^k, T^k]\,[V^n] = 0$ in $\Omega_{n+k}((Z_p)^k)$. Given an Ω-module M, the *annihilator* of $\gamma \in M$ is the set of $[M^n]$ with $\gamma \cdot [M^n] = 0$.

(2.3) *$SF((Z_p)^k)$ is contained in the annihilator $A(\gamma)$ of the torus $\gamma = [(Z_p)^k, T^k] \in \Omega_k((Z_p)^k)$.*

We conjecture that actually $SF((Z_p)^k) = A(\gamma)$.

It is not difficult to show that $A(\gamma)$ is contained in the ideal $I(p)$ of cobordism classes all of whose Pontryagin numbers are divisible by p. We have also settled the structure of $I(p)$.

We go on now to a case in which we are able to give definitive results, namely $SF(Z_p \times Z_p)$. We must first make some remarks concerning maps $T: V^n \to V^n$ of odd prime period p. Let F^m denote a component of the fixed point set F. We must know the structure of the normal bundle $\xi: E \to F^m$. It is seen that the differential dT acts on E, yielding a linear map of period p of each fiber into itself, leaving only the zero vector fixed. We analyze the structure of all such bundles. In particular for each complex vector space bundle $\xi_k: E_k \to F^m$, we get such a map as $v \to \varrho^k \cdot v$ where $\varrho = \exp\,(2\pi i/p)$. We show that ξ splits into a Whitney sum $\xi_1 \oplus \ldots \oplus \xi_{p-1/2}$ of such bundles. That is, ξ can be considered a $U(n_1) \times \ldots \times U(n_{p-1/2})$-bundle for appropriate $n_1, \ldots, n_{p-1/2}$.

(2.4) *$SF(Z_p \times Z_p)$, p an odd prime. is the ideal of Ω generated by $[Y^0]$ and $[Y^{2p-2}]$ where $Y^0 = p$ points and $Y^{2p-2} = P_{p-1}(C)$.*

Outline of proof. Suppose $Z_p \times Z_p$ acts without stationary points on V^n. Denote the generators of $Z_p \times Z_p$ by T, S. There is the free resolution V^n, V^{n+1}, V^{n+2} of the action. The proof uses an analysis of V^{n+1}. Recall that V^{n+1} is a sphere bundle over $F = F(T, V^n)$. Checking back, this is the sphere bundle of the Whitney sum ξ' of a trivial plane bundle with the normal bundle $\xi: E \to F$. Over each component F^m of F, ξ' is a $(U(n_1) \times \ldots \times U(n_{p-1/2})$-bundle.

There is a natural cobordism theory for $U(n_1) \times \ldots \times U(n_{p-1/2})$-bundles, leading to cobordism classes $[\xi']$ of bundles. We next show that $[\xi']$ is divisible by p. There is used here the diagram

$$\begin{array}{ccc} V^{n+1} & \xrightarrow{S} & V^{n+1} \\ \downarrow & & \downarrow \\ F(T, V^n) & \xrightarrow{S} & F(T, V^n), \end{array}$$

an important point being that S acts without fixed points on $F(T, V^n)$. There is $\xi'/S: E'/S \to F/S$. A suitable generalization of the argument of (2.1) shows that $[\xi'] = p[\xi'/S]$, and hence $[\xi']$ is divisible by p.

There is the map $T: V^{n+1} \to V^{n+1}$ acting without fixed points. We thus have an element $[Z_p, V^{n+1}] \in \Omega_{n+1}(Z_p)$. One next has to convert the statement that $[\xi']$ is divisible by p into the statement that $[Z_p, V^{n+1}]$ is divisible by p in $\Omega_{n+1}(Z_p)$. Now

$$[Z_p, V^{n+1}] = [Z_p, S^1]\,[V^n]$$

by (2.2). We are thus reduced to a question purely about the module $\Omega_*(Z_p)$. For what $[V^n]$ is $\alpha_1 \cdot [V^n]$ divisible by p? Fortunately this is a problem we can solve; the set of all such $[V^n]$ is the ideal generated by p and $[P_{p-1}(C)]$.

This illustrates a technique that we use often. We reduce a geometric question to a question about the module $\Omega_*(Z_p)$, and try to solve it there. We give now the last example of a complete computation for $SF(G)$. It uses somewhat similar methods.

(2.5) *The ideal* $SF(Z_3 \times Z_3 \times Z_3)$ *is the ideal generated by* $[Y^0], [Y^4], [Y^{16}]$, *where* $Y^0 = 3$ *points,* $Y^4 = P_2(C)$, $Y^{16} =$ *Cayley plane.*

There is a last computation, almost complete.

(2.6) *The ideal* $SF(T^k)$, T^k *a toral group, is contained in the torsion subgroup of* Ω.

This follows by showing that $SF(T^k) \subset \bigcap_{p \text{ large}} SF((Z_p)^k)$. It then follows from our previous results that every element of $SF(T^k)$ has all its Pontryagin numbers zero. The result then follows.

REFERENCES

[1]. Borel, A., Sous-groupes commutatifs et torsion des groupes de Lie compacts connexes. *Tôhoku Math. J.*, 13 (1961), 216–240.

[2]. Borel, A., *et al.*, Seminar on transformation groups. *Ann. Study*, 46. Princeton, 1960.

[3]. Conner, P. E., & Floyd, E. E., Differentiable periodic maps. *Bull. Amer. Math. Soc.*, 68 (1961), 76–86.

[4]. Milnor, J. W., On the cobordism ring Ω_* and a complex analogue. *Amer. J. Math.*, 82 (1960), 505–521.

[5]. Thom, R., Quelques propriétés globales des variétés differentiables, *Comment. Math. Helv.*, 29 (1954), 17–86.

[6]. Wall, C. T. C., Determination of the cobordism ring. *Ann. Math.*, 72, (1960), 292–311.

ON SOME NUMERICAL HOMOTOPY INVARIANTS

By TUDOR GANEA

The purpose of this paper is to present a series of comparatively recent results concerning Lusternik-Schnirelmann category and several related invariants.

1.

The following definition is essentially due to Lusternik and Schnirelmann [17]:

DEFINITION 1.1. *The category,* cat X, *of a topological space* X *is the least integer* $k \geqslant 0$ *such that* X *may be covered by* $k+1$ *open subsets which are contractible in* X; *if no such integer exists,* cat $X = \infty$.

Let X^k be the k-fold Cartesian power of a topological space X with base-point $*$, let $T = T(X,k)$ denote the subset of all points $(x_1, \ldots, x_k) \in X^k$ such that $x_i = *$ for some i $(1 \leqslant i \leqslant k)$, let $j: T \to X^k$ denote the inclusion map, and let $X^{(k)}$ and $q: X^k \to X^{(k)}$ denote the identification space and identification map resulting by shrinking to a point the subset T of X^k. It has been noticed by G. W. Whitehead [19] that, provided X is a connected CW-complex, cat $X < k$ if and only if there is a map $\Phi: X \to T(X,k)$ such that $j \circ \Phi \simeq \Delta$, where $\Delta: X \to X^k$ is the diagonal map. Moreover, Berstein and Hilton [5] have introduced the following.

DEFINITION 1.2. *The weak category,* wcat X, *of a based topological space* X *is the least integer* $k \geqslant 0$ *such that the composition* $X \xrightarrow{\Delta} X^{k+1} \xrightarrow{q} X^{(k+1)}$ *is null-homotopic rel. base-point; if no such integer exists,* wcat $X = \infty$.

Let ΣX denote the reduced suspension over the based space X and let

$$\sigma: \Sigma X \to \Sigma X \vee \Sigma X \quad \text{and} \quad \tau: \Sigma X \to \Sigma X$$

denote the maps which result, respectively, by pinching to a point the equator of ΣX and by reflecting ΣX in its equator; as shown by Eckmann and Hilton [6], these two maps convert ΣX into an H'-space, the dual of an associative H-space with inversion. The basic co-commutator map ψ is the composition

$$\Sigma X \xrightarrow{\sigma} \Sigma X \vee \Sigma X \xrightarrow{\sigma \vee \sigma} {}^2\Sigma X \vee {}^2\Sigma X \xrightarrow{{}^2 1 \vee {}^2 \tau} {}^2\Sigma X \vee {}^2\Sigma X \xrightarrow{\nabla} {}^2\Sigma X,$$

where ${}^kA$ and ${}^kf$ denote the k-fold wedge of spaces and maps, 1 always stands for the identity map, and ∇ is the folding map given by $\nabla(a,*) = \nabla(*,a) = a$. The co-commutator map ψ_1 of weight 1 is the identity map of ΣX; the co-commutator map ψ_{k+1} of weight $k+1$ is the composition

$$\Sigma X \xrightarrow{\psi} \Sigma X \vee \Sigma X \xrightarrow{\psi_k \vee 1} {}^k\Sigma X \vee \Sigma X = {}^{k+1}\Sigma X$$

in which ψ_k is the co-commutator map of weight $k \geqslant 1$ [2]. The following definition may be motivated by the next proposition, where $\pi(A, B)$ denotes the set of based homotopy classes of maps $A \to B$, which is a group if A is a suspension or if B is a loop space, and where nil π is the classical nilpotency class of the abstract group π.

DEFINITION 1.3. *The co-nilpotency class,* conil X, *of a based topological space X is the least integer $k \geqslant 0$ such that the co-commutator map ψ_{k+1} is nullhomotopic rel. base-point; if no such integer exists,* conil $X = \infty$ [2].

PROPOSITION 1.4. conil $X = \sup \operatorname{nil} \pi(X, \Omega Y)$ *where Y ranges over all based topological spaces* [2].

Let us denote by $\cup$-long X the least integer $k \geqslant 0$ such that, for any commutative coefficient ring, the cup product of any $k+1$ singular cohomology classes of positive dimension vanishes in X; if no such integer exists, $\cup$-long $X = \infty$. The numerical invariants we have defined are related by inequalities as shown [2, 5, 14] in the

THEOREM 1.5. *If X is a connected CW-comples, then*

$$\cup\text{-long } X \leqslant \operatorname{conil} X \leqslant \operatorname{wcat} X \leqslant \operatorname{cat} X.$$

Under the process of attaching a cone to X each of the preceding invariants will increase by at most one unit [2, 5]. Berstein and Hilton [5] have given conditions in terms of the primitivity of the attaching map under which the weak category or the category actually increases by one unit when a cone is attached. Their technique yields the following examples which show that strict inequalities may occur in 1.5.

1.6. Let X result by attaching a 5-cell to S^2 by means of a map in the class generating $\pi_4(S^2)$; then, $\cup$-long $X = 1$ and conil $X = 2$ [5, 3].

1.7. Let X result by attaching to S^2 an 8-cell by means of a map in the class $\{\eta \circ \omega \circ \eta\}$, where $\{\eta\}$ generates $\pi_{n+1}(S^n)$ for $n = 6, 2$, and $\{\omega\}$ generates $\pi_6(S^3)$; then, conil $X = 1$ and wcat $X = 2$ [14].

1.8. Let X result by attaching to $S^3 \vee S^3$ an 8-cell by means of a map in the class of the triple Whitehead product $[\iota_1, [\iota_1, \iota_2]]$, where $\iota_1, \iota_2 : S^3 \to S^3 \vee S^3$ are the classes of the left- and right-hand embeddings; then, wcat $X = 1$ and cat $X = 2$ [5].

Nevertheless, one has [3]:

THEOREM 1.9. *Let X be an $(n-1)$-connected* CW-*complex and suppose* conil $X < k$ $(n \geqslant 1, k \geqslant 1)$. *If* dim $X \leqslant 2kn - 2$, *then* wcat $X < k$; *if* dim $X \leqslant (k+1)n - 2$, *then* cat $X < k$.

One has conil $X < 2$ if and only if the suspension ΣX is homotopy-commutative in the sense that $\gamma \circ \sigma \simeq \sigma$, where $\gamma : \Sigma X \vee \Sigma X \to \Sigma X \vee \Sigma X$ switches the two summands. Therefore, 1.9 yields the following corollary which may be looked at as the dual of a previous result by Sugawara [18, Th. 8.1]:

COROLLARY 1.10. *Let X be an $(n-1)$-connected CW-complex of dimension $\leqslant 3n-2$ $((n \geqslant 1)$. If the suspension ΣX is homotopy-commutative, then* cat $X \leqslant 1$ [3].

In the proofs of 1.5 and 1.9 the following relation between conilpotency and weak category may be used:

PROPOSITION 1.11. *Let X be a connected* CW-*complex. Then,* conil $X<k$ *if and only if the composition* $\Sigma X \xrightarrow{\Sigma\Delta} \Sigma(X^k) \xrightarrow{\Sigma q} \Sigma(X^{(k)})$ *is nullhomotopic.*

We shall now record two results which provide upper bounds for category:

THEOREM 1.12. *If X is a $(p-1)$-connected $(p\geqslant 2)$* CW-*complex such that $H_r(X)$ is free and $H_q(X)=0$ for $q>r$, then* cat $X\leqslant r/p$.

THEOREM 1.13. *Let X be a $(p-1)$-connected $(p\geqslant 2)$* CW-*complex and let $E=\{q>0\,|\,H_q(X)\neq 0\}$. If the set E is contained in the union of k closed linear intervals, each of length $p-2$, then* cat $X\leqslant k$; *if $H_r(X)$ is free, where $r=\max E$, then the last interval may have length $p-1$.*

Theorem 1.12 has first been proved by D. P. Grossman [15] by means of elaborate geometric arguments; 1.13 generalizes a result by Eckmann and Hilton [7] concerning the case $p=2$. Neither of these two results is a particular case of the other, but they may both be derived [12] from a general lemma on induced co-fibrations using two different filtrations of the complex X.

Let now π be an abstract group and let alg dim π stand for the least integer $k\geqslant 0$ such that $H^q(\pi,A)=0$ for all $q>k$ and all π-modules A; if no such integer exists, alg dim $\pi=\infty$. Let $K(\pi,1)$ denote any aspherical CW-complex with π as fundamental group. The following result is due to Eilenberg and the author [8]:

THEOREM 1.14. cat $K(\pi,1)=$alg dim π *with the possible exception of the case* cat $K(\pi,1)=2$ and alg dim $\pi=1$.

The exceptional case is equivalent to the existence of a group π which fails to be free but has the property that each exact sequence $0\to A\to E\to\pi\to 1$ where A is Abelian splits; it is as yet unknown whether such groups exist. The proof of 1.14 is based on an obstruction argument and the following lemma [9] which, upon identifying the "Deckbewegungsgruppe" $G(\tilde{E},p)$ of the simply connected covering space $(\tilde{E},p)$ of E with $\pi_1(E)$, may be regarded as a generalization of classical theorem of Hurewicz [16]:

LEMMA 1.15. *Let $(\tilde{X},f)$ and $(\tilde{Y},g)$ be regular covering spaces of the connected, locally connected spaces X and Y. If X is paracompact and if $\pi_i(\tilde{Y})=0$ for $0\leqslant i<$* cat $(\tilde{X},f)$, *then any homomorphism $H:G(\tilde{X},f)\to G(\tilde{Y},g)$ is induced by a continuous map $h:X\to Y$; here,* cat $(\tilde{X},f)$ *is the least integer $k\geqslant 0$ such that X may be covered by $k+1$ open subsets over each of which $(\tilde{X},f)$ has a cross-section.*

Let us finally note the following [13]:

COROLLARY 1.16. *If X is a closed n-dimensional manifold such that $\pi_1(X)\neq 0$ and $\pi_i(X)=0$ for $2\leqslant i<n$, then* cat $X=n$. *In addition, if $\pi_1(X)$ has a periodic element, then it is necessarily a finite group.*

Further results concerning the category of manifolds have been obtained by I. Berstein [1].

2.

Many of the preceding results may be "dualized" in the sense of Eckmann and Hilton [6]. In this section we shall be exclusively concerned with spaces

with base-point and all maps and homotopies are to preserve base-points. The dual of the Lusternik-Schnirelmann category is the cocategory, cocat X, of an arbitrary topological space X as given by the following inductive [10]:

DEFINITION 2.1. cocat $X=0$ *if and only if X is contractible;* cocat $X\leqslant k+1$ *whenever there exists a fibration $F\to E\to B$ such that the fibre F dominates X and* cocat $E\leqslant k$. *If the phrase* cocat $X\leqslant k$ *is false for any $k\geqslant 0$, we put* cocat $X=\infty$.

Evidently, cocat $X=k$ means that cocat $X\leqslant k$ is true but cocat $X\leqslant k-1$ is false. In the Eckmann-Hilton setting, fibrations are dual to co-fibrations and cocategory is indeed dual to category since, provided X has the based homotopy type of a connected CW-complex, cat X is equal to the invariant obtained by reversing the arrows and replacing in 2.1 the words "fibration" and "fibre" by "co-fibration" and "co-fibre" respectively [10]. The dual of weak category has not yet been defined.

Let ΩX denote the loop space over the based space X and let

$$\mu:\Omega X\times\Omega X\to\Omega X \quad\text{and}\quad \nu:\Omega X\to\Omega X$$

denote the loop multiplication and loop inversion. The basic commutator map ϕ is the composition

$$\Omega X^2\xrightarrow{\Delta}\Omega X^2\times\Omega X^2\xrightarrow{1^2\times\nu^2}\Omega X^2\times\Omega X^2\xrightarrow{\mu\times\mu}\Omega X\times\Omega X\xrightarrow{\mu}\Omega X,$$

where A^k and f^k denote the k-fold Cartesian product of spaces and maps, 1 always stands for the identity map, and Δ is the diagonal map. The commutator map ϕ_1 of weight 1 is the identity map of ΩX; the commutator map ϕ_{k+1} of weight $k+1$ is the composition

$$\Omega X^{k+1}=\Omega X^k\times\Omega X\xrightarrow{\phi_k\times 1}\Omega X\times\Omega X\xrightarrow{\phi}\Omega X$$

in which ϕ_k is the commutator map of weight $k\geqslant 1$ [2].

DEFINITION 2.2. *The nilpotency class,* nil X, *of a based topological space X is the least integer $k\geqslant 0$ such that the commutator map ϕ_{k+1} is nullhomotopic rel. base-point; if no such integer exists,* nil $X=\infty$ [2].

PROPOSITION 2.3. nil $X=\sup$ nil $\pi(\Sigma Y,X)$ *where Y ranges over all based topological spaces* [2].

Let us denote by W-long X the least integer $k\geqslant 0$ such that any $(k+1)$-fold Whitehead product $[[\alpha_1,\ldots,\alpha_k],\alpha_{k+1}]$ with $\alpha_i\in\pi_{q_i}(X)$, $q_i\geqslant 1$, vanishes; if no such integer exists, W-long $X=\infty$. Then [2, 10]:

THEOREM 2.4. W-long $X\leqslant$ nil $X\leqslant$ cocat X.

The CW-complex X obtained by attaching cells to the complex projective plane so as to kill its homotopy groups in dimensions $\geqslant 6$ satisfies nil $X=1$ and cocat $X=2$ [2]; spaces such that W-long $X=1$ and nil $X=2$ are also available. The dual of Theorem 1.9 above has not yet been proved in full generality; the only known result in this line is the previously quoted theorem of Sugawara [18, Th. 8.1] according to which an $(n-1)$-connected CW-complex X such that $\pi_q(X)=0$ for $q>3n-2$ is an H-space provided nil $X=1$.

The connection between H-spaces and cocategory is given by the following [13]

PROPOSITION 2.5. *If* cocat $X \leqslant 1$, *then X has a continuous multiplication with two-sided homotopy unit; the converse is true provided X is a connected countable* CW-*complex.*

The dual of 1.13 reads [11]:

THEOREM 2.6. *Let X be a $(p-1)$-connected $(p \geqslant 2)$* CW-*complex and let $E = \{q > 0 \mid \pi_q(X) \neq 0\}$. If the set E is contained in the union of k closed linear intervals, each of length $p-2$, then* cocat $X \leqslant k$.

Theorems 2.4 and 2.6 may be used to construct spaces of arbitrarily given cocategory. As a particular case of 2.4 one has nil $\pi_1(X) \leqslant$ nil X; for aspherical spaces a more precise result is available [10], which again yields spaces of arbitrarily given cocategory:

THEOREM 2.7. nil $\pi =$ nil $K(\pi, 1) =$ cocat $K(\pi, 1)$.

The process of killing off homotopy groups does not increase cocategory. In fact, if (X, n) denotes the $(n-1)$-connective space over X and if (n, X) is obtained by attaching cells to X so as to kill off its homotopy groups in dimensions $\geqslant n$, then

THEOREM 2.8. *If X is a connected* CW-*complex, then* cocat $(X, n) \leqslant$ cocat X *and* cocat $(n, X) \leqslant$ cocat X *for any* $n \geqslant 1$ [11].

COROLLARY 2.9. *The simply connected covering $\tilde{X}$ of a connected* CW-*complex X satisfies* cocat $\tilde{X} \leqslant$ cocat X.

Let Y^X denote the compact-open topologized space of all base-points preserving continuous maps $X \to Y$ with the constant map as base-point.

THEOREM 2.10. *If X is a connected* CW-*complex, then* cocat $Y^X \leqslant$ cat X; *if X is a Hausdorff locally compact space, then* cocat $Y^X \leqslant$ cocat Y [10].

Let us finally mention that satisfactory definitions of the category and cocategory of continuous maps have also been found and many of the above results may thus be generalized [4, 10].

REFERENCES

[1]. BERSTEIN, I., Sur la catégorie de Lusternik-Schnirelmann. *C. R. Acad. Sci. Paris*, 246 (1958), 362–364.

[2]. BERSTEIN, I. & GANEA, T., Homotopical nilpotency. *Illinois J. Math.*, 5 (1961), 99–130.

[3]. —— On the homotopy-commutativity of suspensions. *ibid.* 6 (1962). 336–340.

[4]. —— The category of a map and of a cohomology class. *Fund. Math.*, 50 (1962), 265–279.

[5]. BERSTEIN, I. & HILTON, P. J., Category and generalized Hopf invariants. *Illinois J. Math.*, 4 (1960), 437–451.

[6]. ECKMANN, B. & HILTON, P. J., Groupes d'homotopie et dualité. *C. R. Acad. Sci. Paris*, 246 (1958), 2444–47, 2555–58, 2991–93.

[7]. —— On the homology and homotopy decomposition of continuous maps. *Proc. Nat. Acad. Sci. U.S.A.*, 45 (1959), 372–375.

[8]. EILENBERG, S. & GANEA, T. ,On the Lusternik-Schnirelmann category of abstract groups. *Ann. Math.*, 65 (1957), 517–518.

[9]. GANEA, T., Catégorie 1-dimensionnelle et homomorphismes de groupes fondamentaux. *C. R. Acad. Sci. Paris*, 242 (1956), 1407–10.

[10]. —— Lusternik-Schnirelmann category and cocategory. *Proc. London Math. Soc.*, 10 (1960), 623–639.

[11]. —— Fibrations and cocategory. *Comment. Math. Helv.*, 35 (1961), 15–24.

[12]. —— Upper bounds for the Lusternik-Schnirelmann category, *Dokl. Akad. Nauk S.S.S.R.*, 136 (1961), 1273–76. (In Russian.)

[13]. —— Sur quelques invariants numériques du type d'homotopie. *Séminaire de Topologie et de Géometrie Différentielle, C. Ehresmann.* Paris (1962).

[14]. GANEA, T., HILTON, P. J. & PETERSON, F. P., On the homotopy-commutativity of loop spaces and suspensions. *Topology*, 1 (1962), 133–141.

[15]. GROSSMAN, D. P., An estimation of the category of Lusternik-Schnirelmann. *C. R.* (*Dokl.*) *Acad. Sci. U.R.S.S.* (*NS*), 54 (1946), 109–112.

[16]. HUREWICZ, W., Beiträge zur Topologie der Deformationen. IV. *Proc. Acad. Sci. Amsterdam*, 39 (1936), 215–224.

[17]. LUSTERNIK, L. & SCHNIRELMANN, L., *Méthodes topologiques dans les problèmes variationnels.* Herman et Cie, Paris, 1934.

[18]. SUGAWARA, M., On the homotopy-commutativity of groups and loop spaces. *Mem. Coll. Sci. Univ. Kyoto*, 33 (1960) 257–269.

[19]. WHITEHEAD, G. W., The homology suspension. *Coll. Topologie Algébrique.* Louvain, 1956, 89–95.

ALLGEMEINE STETIGKEITSSTRUKTUREN

Von MIROSLAV KATĚTOV

Es ist für die Entwicklung der allgemeinen Topologie in den letzten Jahren kennzeichnend, dass neben den klassischen Strukturen, nämlich topologischen und uniformen Räumen, verschiedene andere Typen von Räumen untersucht werden, manchmal an sich, manchmal als Hilfsmittel zur Erforschung von klassischen Gebilden. Es seien nur erwähnt: Berührungsräume (die Theorie dieser Räume wurde systematisch von J. Smirnov ausgebaut; für Literaturangeben siehe z. B. den Bericht von M. Katětov [11]); syntopogene Räume (A. Császár [4]); Kontingenz- oder Angrenzungsräume (V. M. Ivanova, A. A. Ivanov [9]); Untersuchung von Lipschitz-invarianten Eigenschaften der metrischen Räume (M. Katětov [12], vgl. auch N. Aronszajn, P. Panitchpakdi [3] und L. Géher [7]).

Den Gegenstand der allgemeinen Topologie bilden schon längst nicht nur topologische Räume, sondern vielmehr verschiedene Typen von Räumen, ihre Zusammenhänge und mannigfaltige Prozesse der Erzeugung von Räumen. Die unmittelbare Beziehung zu den anschaulichen geometrischen Gebilden tritt oft in den Hintergrund, um so enger aber werden die Beziehungen zur Funktionalanalysis und zur topologischen Algebra.

In diesem Bericht wird, ohne auf einzelne Ergebnisse einzugehen und manchmal mit Verzicht auf präzise Formulierung, ein Versuch gemacht, einige neuere Begriffsbildungen und Fragestellungen der allgemeinen Topologie von einem einheitlichen Gesichtspunkt zu beleuchten.

Versucht man die intuitiven Ideen zu beschreiben, die mit jedem mathematischen Gebilde verbunden sind, welches als eine Stetigkeitsstruktur aufgefasst werden kann, so scheint der Gedanke einer „beliebig guten Approximation" auf der Hand zu liegen. Für jedes derartige Gebilde liegt fest, wann gewisse Objekte x_α durch andere Objekte y_α (α veränderlich) beliebig gut approximiert werden. So wird eine Topologie bestimmt durch Angaben über die Approximation (im Sinne $x \in \overline{M}$) von jedem einzelnen Punkte x durch Punkte $y \in M$, während es sich bei einer gleichmässigen Struktur um die Approximation von einem „veränderlichen" x_α durch y_α handelt. Im allgemeinen geht es jedoch nicht um die Approximation von Punkten, sondern von gewissen Objekten, die auf eine natürliche gesetzmässige Weise mit der betrachteten Grundmenge verbunden sind, z. B. um Approximation von Untermengen durch Untermengen.

Der Gedanke einer unendlich engen Approximation hängt freilich mit der Idee von „beliebig kleinen Objekten" zusammen. Werden nämlich Objekte ξ_α beliebig gut durch Objekte η_α approximiert, so werden die „Unterschiede" von ξ_α, η_α beliebig klein, und umgekehrt. Es erscheint daher intuitiv angemessen, den Begriff einer allgemeinen Stetigkeitsstruktur auf den Begriff von beliebig kleinen Objekten zurückzuführen. In diesem Sinne kann also die allgemeine Topologie als eine allgemeine Lehre von Unendlichkleinen aufgefasst werden.

Wir gehen jetzt zu den exakten Definitionen über. Es sei Z eine Menge. Es sei Γ ein System von nichtleeren $S \subset Z$ derart, dass (1) mit S enthält Γ

alle S_1 mit $S \subset S_1 \subset Z$, (2) ist $S_1 \cup S_2 \in \Gamma$, so ist $S_1 \in \Gamma$ oder $S_2 \in \Gamma$. Dann soll Γ eine I-Struktur auf Z heissen. Offensichtlich drücken beide Forderungen intuitiv unentbehrliche Eigenschaften des Begriffs eines Systems mit beliebig kleinen Elementen aus (vgl. dazu V. Sandberg [16]).

Sind Γ, Γ' I-Strukturen auf Z, so heisst Γ feiner, bzw. gröber als Γ', falls $\Gamma \supset \Gamma'$, bzw. $\Gamma \subset \Gamma'$. Stetige Abbildungen oder I-Morphismen $f:(Z,\Gamma) \to (Z', \Gamma')$ werden auf eine naheliegende Weise definiert (durch die Bedingung $f\Gamma \subset \Gamma'$).

Es sei nun $\mathfrak{M}$ die Kategorie aller Mengen, $\Phi: \mathfrak{M} \to \mathfrak{M}$ ein kovarianter Funktor. Eine I-Struktur Γ auf ΦX heisst dann eine Φ-Struktur auf X und (X,Γ) heisst ein Φ-Raum. Eine Abbildung $f: X \to Y$ heisst Φ-stetig oder ein Φ-Morphismus, falls die entsprechende Abbildung $\Phi f: \Phi X \to \Phi X$ ein I-Morphismus ist.

Es sei Q der Funktor, der jeder Menge X ihr „Quadrat" $X \times X$ zuordnet. Wir betrachten nur solche Q-Strukturen Γ, welche jedes $\{(x,x)\}$, $x \in X$, enthalten. Diejenigen Q-Räume, für welche es zu jeden $S \in \Gamma$ eine Menge $T \in \Gamma$, $T \subset S$, von der Gestalt $T = (a) \times B$ gibt, entsprechen dann genau den topologischen Räumen (die Abgeschlossenheit der Abschliessung wird nicht verlangt).

Ein zweites Beispiel: Diejenigen Q-Räume, für welche es zu jedem $S \subset Z$, $S \notin \Gamma$, eine Menge $T \supset S$, $T \notin \Gamma$, von der Gestalt $T = \bigcup_{k=1}^{n} A_k \times B_k$ gibt, ergeben Berührungsräume (δ-Räume) in einer verallgemeinerten Fassung (vgl. u. a. C. H. Dowker [5], V. M. Ivanova, A. A. Ivanov [9]).

Wird schliesslich vorausgesetzt, dass das System von denjenigen $U \subset X \times X$, für welche $X \times X - U$ nicht in Γ liegt, die wohlbekannten Axiome für Umkreisungen der Diagonale erfüllt, so erhält man die uniformen Räume im klassischen Sinne.

Es seien noch die P-Strukturen erwähnt: P ordnet jeder Menge X ihre Potenzmenge pot X zu; diese Strukturen sind mit den von A. Ivanov [8] betrachteten „gleichmässigen Strukturen" im Wesentlichen äquivalent. Jeder Q-Struktur Γ wird eine P-Struktur Γ^* zugeordnet: $\mathfrak{M} \in \Gamma^*$ dann und nur dann, wenn $\{(X_M, Y_M)\} \in \Gamma$ für jede Wahl von $X_M \in M$, $Y_M \in M$, wo $M \in \mathfrak{M}$; jedoch lässt sich nicht jede P-Struktur so erzeugen.

Eine allgemeine Theorie der Φ-Strukturen existiert freilich noch nicht; insbesondere ist es eine offene Frage, ob und wie die Kompaktheit, Vollständigkeit usw. definiert werden können. Einige Definitionen und Sätze können jedoch ganz allgemein formuliert werden.

Ist X eine Menge und sind f_α Abbildungen von Φ-Räumen Z_α in X, so gibt es eine (eindeutig bestimmte) feinste Φ-Struktur auf X, für welche alle f_α Φ-stetig sind; wir wollen sagen, dass sie durch die Abbildungen f_α induktiv erzeugt wird. Analog wird die projektive Erzeugung durch Abbildungen $g_\alpha: X \to Y_\alpha$ definiert. Daraus ergibt sich zugleich eine allgemeine Definition des Unterraumes, des Quotienten, der Summe und des Produktes von Räumen u. a. Es ist jedoch zu beachten, dass die allgemeine Definition des kartesischen Produktes für einige Raumtypen nicht ganz zweckmässig ist.

Ich betrachte jetzt die induktive und projektive Erzeugung von Q-Räumen. Im Falle einer Erzeugenden handelt es sich dann um den Quotientenraum Z/f, bzw. (im projektiven Fall) um eine Einbettung; übrigens lässt sich der allgemeine Fall darauf reduzieren (auf den Quotienten der Summe, bzw. auf die Einbettung in das Produkt).

Es zeigt sich, dass jeder topologische Raum als ein Quotient von einem

parakompakten nulldimensionalen Raum dargestellt werden kann. Es ist zu beachten, dass bei der hier zugrundeliegenden Definition des Quotienten im allgemeinen die Beziehung $\bar{\bar{A}} = \bar{A}$ nicht gilt, dafür aber ist der Übergang zu Unterräumen in einem bestimmten Sinne mit der Bildung von Quotienten vertauschbar. Es zeigt sich weiter, dass sich jeder Q-Raum (mit einer symmetrischen Struktur) als ein Quotient eines uniformen Raumes darstellen lässt, welcher sogar topologisch diskret gewählt werden kann; die Bedingung der Symmetrie kann fortgelassen werden, wenn verallgemeinerte uniforme („ordoforme") Räume im Sinne von D. Tamari [17] betrachtet werden.

Über die Einbettung ist für allgemeine Q-Strukturen nichts bekannt; ein beliebiger topologischer Raum kann jedoch in das Produkt von genügend vielen Exemplaren eines bestimmten aus drei Punkten bestehenden Raumes eingebettet werden.

In dem Falle, wo mehrere Erzeugende oder zusätzliche Bedingungen auftreten, ist die Lage weit mehr kompliziert. Für uniforme Räume, geschweige denn für allgemeine Q-Strukturen, ist wenig bekannt, für topologische Räume sind jedoch zahlreiche Ergebnisse vorhanden.

Es wurden u. a. untersucht: (1) der Fall einer gerichteten Menge von Abbildungen, was auf den projektiven bzw. induktiven Grenzwert führt; (2) der „gleichmässige" Fall, der als Erzeugung von den assoziierten P-Strukturen aufgefasst werden kann; (3) für eine induktive Erzeugende: Erzeugung durch offene, abgeschlossene u. ä. Abbildungen; (4) ebenfalls für eine Erzeugende: Eigenschaften von Urbildern $f^{-1}(x)$, bzw. Lageeigenschaften von $g(X)$. Für eine Darstellung der betreffenden Ergebnisse sei auf zwei Artikel von P. S. Alexandroff [1, 2] verwiesen; nur als Illustration seien einige Sätze angeführt (vgl. u. a. V. Ponomarev [14, 15], V. Trnková [17]).

Metrisierbare (bzw. separable metrisierbare) Räume erzeugen projektiv (a) ohne zusätzliche Bedingungen — alle vollständig regulären Räume, (b) als Grenzwerte — alle topologisch vollständigen (bzw. sog. reell-kompakten) Räume, (c) gleichmässig — alle parakompakten (bzw. Lindelöfschen) Räume. Diskrete Räume erzeugen gleichmässig projektiv alle parakompakten nulldimensionalen Räume, die wieder durch perfekte Abbildungen alle parakompakten Räume erzeugen.

Es sei noch folgendes bemerkt: ein induktives System $f_\alpha : Z_\alpha \to X$ sei F-, bzw. G-erblich genannt, falls seine Einengung $f'_\alpha : Z'_\alpha \to X$ auf abgeschlossene, bzw. offene Z_α auf der Menge $\cup\, f'_\alpha(Z'_\alpha)$ die Topologie eines Unterraumes von X erzeugt. Es zeigt sich, dass (unter gewissen Voraussetzungen) $f : Z \to X$ dann und nur dann abgeschlossen, bzw. offen ist, wenn sie die Gestalt $Z \to Z/f$ mit einem F-, bzw. G-erblichen f hat. Auf eine verwandte Weise lassen sich übrigens auch perfekte Abbildungen charakterisieren. (Eine Abbildung $Z \to Z/f$ ist dann und nur dann perfekt, d. h. abgeschlossen und kompakt, wenn eine gewisse Vertauschungsrelation gilt, nämlich wenn ein kompakter Raum bZ und eine Abbildung f existieren derart, dass Z/f als ein Unterraum von bZ/f betrachtet werden kann).

Was die uniformen Räume betrifft, ist der Fall einer „Selbstinduktion" erwähnenswert: es sei (X, μ) ein uniformer Raum; für jede gleichmässige Überdeckung $\mathcal{U} = \{U_\alpha\}$ erzeugen die Abbildungen $U_\alpha \to X$ eine Q-Struktur $\mu(\mathcal{U})$ und alle $\mu(\mathcal{U})$ erzeugen (projektiv) eine uniforme Struktur $\lambda\mu$. Es wurden untersucht (J. R. Isbell, S. Ginsburg [10]) der Funktor λ und die Räume (genannt lokal fein), für welche $\lambda\mu = \mu$ gilt.

Ich füge jetzt einige Bemerkungen über QP- und P-Strukturen an (der Funktor QP ordnet jeder Menge X die Menge pot $X \times$ pot X zu).

Wie bekannt, werden syntopogene Strukturen (A. Császár [4]) als gewisse Systeme S von Relationen σ zwischen Untermengen einer gegebenen Menge X definiert, welche gewisse einfache Bedingungen erfüllen. Mit einem syntopogenen Raum (Y, S) wird eine QP-Struktur Γ wie folgt assoziiert: ein system $\mathfrak{M}$ von Paaren (A, B), $A \subset X$, $B \subset X$ soll dann und nur dann in liegen, wenn es für jedes $\sigma \in S$ ein Paar $(A, B) \in \mathfrak{M}$ gibt derart, dass A non σ $(X - B)$. Jede QP-Struktur, die gewisse natürliche Bedingungen erfüllt, lässt sich auf diese Weise erzeugen, wobei zwei syntopogene Strukturen dieselbe QP-Struktur dann und nur dann erzeugen, wenn sie äquivalent sind.

P-Strukturen finden eine interessante Anwendung in der sog. deskriptiven Theorie der topologischen Räume (Z. Frolík [6]). Eine Menge X sei mit einer P-Struktur Γ und einer Topologie τ versehen. Dann heisst Γ vollständig in Bezug auf τ, falls für jedes zentrierte $\mathfrak{M} \in \Gamma$ der Durchschnitt von Mengen $\overline{M}$, $M \in \mathfrak{M}$, nichtleer ist.

„Analytische" (Suslinsche) Räume können nun als diejenigen topologischen Räume definiert werden, die eine vollständige P-Struktur besitzen, die auf eine spezielle Weise durch ein System $\{M_s\}$, wo s endliche Folgen natürlicher Zahlen durchläuft, erzeugt wird; auf dieser Grundlage kann eine systematische „deskriptive" Theorie von topologischen Räumen ausgebaut werden; u. a. entstehen durch die Suslinsche Operation, durchgeführt an kompakten Räumen, genau alle „analytischen Räume"; die Klasse von diesen Räumen entsteht durch stetige Abbildungen von denjenigen Räumen, welche in einem kompakten Raum Borelsche Mengen (in einem verschärften Sinne) sind.

Ich gehe jetzt zu Stetigkeitsstrukturen von einer wesentlich anderen, bisher wenig untersuchten Art über. Es sei Λ der Funktor, der jeder Menge X den Modul ΛX von linearen Kombinationen $\Sigma \lambda_i x_i$, λ_i Zahlen, $x_i \in X$, zuordnet. Eine Λ-Struktur (M. Katětov [13]) ist dann eine I-Struktur auf dem Modul ΛX, wobei verlangt wird, dass ΛX, mit dieser Struktur versehen, zu einem lokal-konvexen topologischen linearen Raum wird. Eine Λ-Struktur auf X ist also gegeben, sobald es bekannt ist, welche $\Sigma \lambda_i x_i$ als klein gelten; intuitiv kann dies so aufgefasst werden, dass die Menge X mit einer „geometrischen" Struktur versehen ist, welche sich darin äussert, dass sich gewisse (in endlichvielen Punkten konzentrierte) Ladungen auf X voneinander wenig unterscheiden. Offensichtlich ist eine Λ-Struktur auf X vollkommen bestimmt durch die Angabe einer Klasse von Abbildungen (welche Λ-morph genannt werden) von X in lokal-konvexe topologische lineare Räume.

Ist X ein Λ-Raum, so kann die vollständige Hülle $\pi \Lambda X$ von ΛX gebildet werden. Die Elemente von $\pi \Lambda X$ können als Distributionen (in einem allgemeinen Sinne) auf X aufgefasst werden; für geeignete Strukturen stimmen sie mit den Distributionen (verallgemeinerten Funktionen) im bekannten Sinne überein.

Offensichtlich erzeugt jede Λ-Struktur auf X eine uniforme und daher auch eine Berührungs- und topologische (vollständig reguläre) Struktur. Umgekehrt kann jede solche Struktur durch eine Λ-Struktur erzeugt werden; unter solchen Strukturen gibt es sogar immer eine feinste, falls es sich um die Erzeugung von einer uniformen Struktur oder einer Topologie handelt (für eine Berührungsstruktur gilt dies nicht). Ein uniformer oder

ein Berührungsraum ist dann und nur dann beschränkt, wenn er für eine geeignete Λ-Struktur beschränkt ist (als eine in ΛX eingebettete Menge); dasselbe gilt analog für totalbeschränkte Räume.

Ich möchte jetzt einige Beispiele anführen. Es sei M ein Modul von Funktionen auf X; werden $f \in M$ zu linearen Formen auf ΛX fortgesetzt, so bestimmen sie eine schwache Topologie auf ΛX, folglich eine Λ-Struktur auf X, welche auch schwach heissen soll. Ist X ein Q-Raum, $M = C(X)$, so erzeugt diese Λ-Struktur die ursprüngliche Topologie von einem vollständig regulären X, im allgemeinen jedoch nicht die ursprüngliche uniforme Struktur.

Es sei X eine komplexe Mannigfaltigkeit, und M bestehe aus allen holomorphen Funktionen auf X. Im allgemeinen wird dann nicht einmal die ursprüngliche Topologie von X reproduziert. Ist es der Fall, so entsteht das Problem, ob die (feinste) Berührungsstruktur reproduziert wird. Schon für die komplexe Ebene ist diese Frage nicht trivial, die Antwort scheint aber in diesem Falle positiv zu sein.

Es sei (X, ϱ) ein metrischer Raum. Es gibt dann Normen ν auf ΛX mit der Eigenschaft $\nu(x-y) = \varrho(x,y)$; unter ihnen gibt es eine, die in einem gewissen Sinne die grösste ist; die durch sie erzeugte Λ-Struktur soll eine Lipschitzsche Λ-Struktur heissen. Zwei solche Strukturen, die den Metriken ϱ, ϱ' entsprechen, sind dann und nur dann gleich, wenn ϱ und ϱ' Lipschitzäquivalent sind. Die Untersuchung von Lipschitzschen Λ-Strukturen ist daher mit dem Studium von Lipschitz-invarianten Eigenschaften von metrischen Räumen gleichbedeutend. Einige einfachste Fragen in dieser Richtung wurden bereits untersucht (M. Katětov [12]); es handelt sich u. a. um die Begriffe des Retraktes, Umgebungsretraktes usw., deren Theorie sich in einigen wesentlichen Zügen von der Theorie der entsprechenden topologischen oder uniformen Begriffen unterscheidet.

Ist ein Λ-Raum $(X, \Gamma_0.)$ gegeben, so können diejenigen Λ-Strukturen Γ betrachtet werden, welche dieselben Λ-morphen Funktionen ergeben (man schreibt dann $\Gamma \sim \Gamma_0.$). Jede solche Struktur kann durch ein geeignetes System von Pseudonormen ν auf X bestimmt werden, und jedes ν wiederum durch ein passendes System von Ungleichungen $|\xi_\alpha| \leqslant c_\alpha$, wo $\xi_\alpha \in \Lambda X$ ($\nu(z)$ wird als die grösste untere Grenze der Zahlen $\Sigma |\lambda_{\alpha(i)}| c_{\alpha(i)}$ für alle $\Sigma \lambda_{\alpha(i)} \xi_{\alpha(i)} = z$ festgesetzt; es wird bei der Wahl von C_α verlangt, dass jedes lineare f mit $|f(\xi_\alpha)| \leqslant c_\alpha$ Λ-morph ist).

Ist etwa X die reelle Gerade mit der durch alle unbeschränkt differenzierbare Funktionen mit kompakten Trägern bestimmten Λ-Struktur, so betrachte man für jede Wahl von c_k, $k = 1, 2, \ldots$ die folgenden Ungleichungen:

$$\left|\frac{1}{h}((x+h)-x)\right| \leqslant c_1, \quad \left|\frac{1}{h^2}((x+2h)-2(x+h)+x)\right| \leqslant c_2, \ldots,$$

x und $h > 0$ beliebig. In diesem Falle besteht $\pi \Lambda X$ genau aus allen Distributionen im gewöhnlichen Sinne.

Die feinste von allen Λ-Strukturen $\Gamma \sim \Gamma_0$ ist oft wenig brauchbar, da sie gewöhnlich auf die diskrete Topologie von X führt. Es ist deshalb zweckmässig, die feinste Λ-Struktur $\Gamma \sim \Gamma_0$ zu suchen, welche die Q-Struktur von X reproduziert. Ist etwa X von vornherein als ein Q-Raum gegeben und ist $M = C(X)$, so wird als $\pi \Lambda X$ der sogenannte freie vollständige lokal-konvexe topologische lineare Raum über X erhalten.

Die Einführung des kartesischen Produkten von Λ-Räumen stösst auf erhebliche Schwierigkeiten (sollte z. B. die allgemeine Definition benützt werden, so würde das Produkt immer ein nichtseparierter Raum sein). Die Lage ist schon für die Lipschitzschen Λ-Räume kompliziert (dies hängt damit zusammen, dass das Produkt von zwei reellwertigen Funktionen, welche die Lipschitz-Bedingung erfüllen, diese Bedingung nicht zu erfüllen braucht).

Folgenden Definition könnte sich vielleicht als zweckmässig erwiesen: Die Topologie von $\Lambda(X \times Y) = X \otimes \Lambda Y$ wird dadurch bestimmt, dass eine lineare Abbildung $\varphi : \Lambda X \otimes \Lambda Y \to Z$ dann und nur dann als stetig gilt, wenn die bilineare Abbildung $(\xi, \eta) \to \varphi(\xi \otimes \eta)$ gleichgradig stetig in $\eta \in \Lambda Y$ für $\xi \in A$ und in $\xi \in \Lambda X$ für $\eta \in B$ ist, wo $A \subset X$, bzw. $B \subset Y$ beliebige beschränkte Mengen sind.

In einigen Fragen, z. B. im Zusammenhang mit differenzierbaren Mannigfaltigkeiten, tritt auch eine ganz andere Art von Produkt auf: sind X, Y Λ-Räume, so wird $X \times Y$ mit der von $(x) \times Y, X \times (y)$ induktiv erzeugten Struktur versehen.

Es sei jetzt X von vornherein mit einer algebraischen Struktur, etwa einer Halbgruppe, ausgestattet. Dann wird ΛX offensichtlich zu einer Algebra, wobei die Abbildung $(\xi, \eta) \to \xi \cdot \eta$ sich unter gewissen Umständen auf alle $\xi, \eta \in \pi \Lambda X$ stetig fortsetzen lässt. In dieser Weise können wohlbekannte Operationen erhalten werden, wie die Faltung von Funktionen u. ä.

Zum Abschluss werde noch eine Art von Φ-strukturen betrachtet. Es sei eine Garbe auf einem topologischen Raum X gegeben: jeder offenen Menge $U \subset X$ sei ein Modul $\mathfrak{F}(U)$ von Funktionen zugeordnet. Für jedes U induziert dann $\mathfrak{F}(U)$ eine Λ-Struktur, wobei für $V \subset U$ eine natürliche stetige lineare Abbildung $\pi \Lambda V \to \pi \Lambda U$ definiert ist. Es wird dadurch eine Art von einer „kontravarianten Garbe" von linearen Räumen erhalten. Die Untersuchung von solchen und ähnlichen Gebilden scheint interessant zu sein; ich möchte jedoch eher hervorheben, dass hier eigentlich ein Speziellfall einer Struktur vorliegt, die wie folgt beschrieben werden kann. Eine Menge T ist gegeben; einige Untermengen U sind mit einer Φ-Struktur (etwa einer Topologie) versehen, wobei für $U_1 \subset U$ die Abbildung $U_1 \to U$ ein Φ-Morphismus ist. Es liegt dann eine Ψ-Struktur auf T vor, wo $\Psi = \Phi \times P$ (d. h. $\Psi X = \Phi X \times \text{pot } X$) ist. Intuitiv handelt es sich um Systeme von „relativ beliebig kleinen" Objekten.

LITERATUR

[1]. Александров, П. С., О некоторых результатах в теории топологических пространств, полученных за последние двадцать пять лет. *Успехи матем. наук*, 15, 2 (92) (1960), 25–95.

[2]. Alexandroff, P., On some results concerning topological spaces and their continuous mappings. *General Topology and its Relations to Modern Analysis and Algebra. Proc. Symp. Prague*, 1961, 41–54.

[3]. Aronszajn, N. & Panitchpakdi, P., Extension of uniformly continuous transformations and hyperconvex metric spaces. *Pacific. J. Math.*, 6 (1956), 405–439.

[4]. Császár, A., *Fondements de la Topologie Générale*. Budapest, 1960.

[5]. Dowker, C. H., Mappings of proximity structures. *General Topology and its Relations to Modern Analysis and Algebra. Proc. Symp. Prague*, 1961, 139–141.

[6]. Frolík, Z., Descriptive theory of sets and spaces. *General Topology and its Relations to Modern Analysis and Algebra. Proc. Symp. Prague*, 1961, 157–173.

[7]. Géher, L., Über Fortsetzungs- und Approximationsprobleme für stetige Abbildungen von metrischen Räumen. *Acta Sci. Math. Szeged*, 20 (1959), 48–66.

[8]. Иванов, А. А., Отношения смежности в топологических проотранствах. *Докл. АН С.С.С.Р.*, 128 (1959), 33–36.

[9]. Иванова, В. М., и Иванов, А. А., Пространства смежности и бикомпактные расширения топологических пространств. *Докл. АН С.С.С.Р.*, 127 (1959), 30–32.

[10]. Isbell, J. R. & Ginsburg, S., Some operators on uniform spaces. *Trans. Amer. Math. Soc.*, 93 (1958), 145–168.

[11]. Katětov, M., Über die Berührungsräume. *Wiss. Z. Humboldt-Univ. Berlin. Math. nat. R.*, 9 (1959/60), 685–691.

[12]. —— О квазиметрических свойствах. (Сообщение на конференции по функциональному анализу, Варшава, 1960 — в печати.)

[13]. —— On a category of spaces. *General Topology and its Relations to Modern Analysis and Algebra. Proc. Symp. Prague*, 1961, 226–229.

[14]. Ponomarev, V., On paracompact spaces and related questions. *General Topology and its Relations to Modern Analysis and Algebra. Proc. Symp. Prague*, 1961, 302–306.

[15]. Пономарев, В., О паракомпактных пространствах и их непрерывных отображениях. *Докл. АН С.С.С.Р.*, 143 (1962), 46–49.

[16]. Сандберг, В. Ю., Новое определение равномерных пространств. *Докл. АН С.С.С.Р.*, 135 (1960), 535–537.

[17]. Tamari, D., On a generalization of uniform structures and spaces. *Bull. Res. Council Israel*, 3 (1954), 417–428.

[18]. Трнкова, В., О замыкании классов пространства по ω-отображениям. (в печати — *Докл. АН С.С.С.Р.*)

DIFFERENTIABLE STRUCTURES ON SPHERES AND HOMOTOPY

By MICHEL A. KERVAIRE

This talk was a report and brief survey of results obtained jointly by J. Milnor and the speaker on the groups of differentiable structures on spheres.

Let $|\theta_n|$ denote the number of differentiable structures on S^n (up to diffeomorphism), then

$$|\theta_{4k-1}| = |\pi_{4k-1}| \cdot 2^{2k-4}(2^{2k-1}-1) \cdot B_k a_k/k,$$

and

$$|\theta_{4k}| = |\pi_{4k}| \cdot a_k/2, \quad |\theta_4| = 1,$$

for $k>1$, where B_k is the kth Bernoulli number, and a_k equals 1 for k even and 2 for k odd. ($B_1=1/6$, $B_2=1/30$, $B_3=1/42$, $B_4=1/30$, $B_5=5/66$, $B_6=691/2730$, etc.)

For all non-negative integers m, one has

$$|\theta_{4m+1}| = |\pi_{4m+1}| \cdot a_m u_m/2$$

and

$$|\theta_{4m+2}| = |\pi_{4m+2}| \cdot u_m/2$$

where a_m is as above, and u_m is either 1 or 2. The only known values of u_m are: $u_0=1$, $u_1=1$, $u_2=2$, $u_3=1$, $u_4=2$. If S^{2m+1} is parallelizable, then $u_m=1$. It is conjectured that the converse also holds, i.e. that $u_m=2$ except for $m=0$, 1 and 3.

In the above formulae $|\pi_n|$ denotes the order of the stable homotopy groups $\pi_n=\pi_{n+N}(S^N)$, $N \geqslant n+2$. ($|\pi_1|=|\pi_2|=2$, $|\pi_3|=24$, $|\pi_4|=|\pi_5|=1$, $|\pi_6|=2$, $|\pi_7|=240$, $|\pi_8|=4$, $|\pi_9|=8$, $|\pi_{10}|=6$, $|\pi_{11}|=504$, $|\pi_{12}|=1$, $|\pi_{13}|=3$, $|\pi_{14}|=4$, $|\pi_{15}|=960$, etc.)

The proofs will be included in two papers: Groups of Homotopy Spheres, I and II, to be published in the *Annals of Mathematics*.

PRODUITS CARTÉSIENS DE VARIÉTÉS DIFFÉRENTIELLES PAR UN DISQUE

Par VALENTIN POÉNARU

1. Introduction

Dans ce qui s'ensuit nous serons principalement concernés avec des problèmes du type suivant :

I. Quelles conditions doit-on imposer aux variétés M_1^n, M_2^n, pour que l'on puisse affirmer que :

$$M_1^n \times D^k = M_2^n \times D^k. \tag{1}$$

Sauf mention contraire toute variété sera supposée de classe C^∞, compacte, éventuellement à bord, égalité signifiera difféomorphisme, D^k le disque k-dimensionnel, E^k l'espace euclidien.

II. Etant donnée une égalité de type (1) qu'est-ce qu'on peut dire de M_1^n, M_2^n ?

Les problèmes plus-haut cités sont « compacts ». On peut évidemment formuler des problèmes « ouverts » analogues, qui s'obtiennent en remplaçant D^k par E^k. Mais l'égalité :

$$M_1^n \times E^k = M_2^n \times E^k \tag{2}$$

est beaucoup plus faible que (1).

En utilisant un théorème de B. Mazur [12] on peut prouver que les espaces lenticulaires $L_{7,1}$ et $L_{7,2}$ qui n'ont pas le même type simple d'homotopie (et qui ne puissent donc satisfaire à aucune égalité de type (1)) satisfont à une relation de type (2) avec $k=4$. C'est ce fait, combiné à $L_{7,1} \times S^3 \neq L_{7,2} \times S^3$(1) qui a permis à J. Milnor [12] d'infirmer le « Hauptvermutung ».

D'autre part (1) n'implique pas nécessairement le difféomorphisme, même si M_1^n, M_2^n sont sans bords (autrement c'est trivial); on peut s'en convaincre en prenant pour $M_1^n = S^n, M_2^n$ certaine sphère de J. Milnor. Mais on peut toujours raisonnablement se demander si deux variétés sans bords, satisfaisant à (1) ne sont pas combinatoirement équivalentes ou tout au moins homéomorphes.

D'autre part (toujours pour les variétés sans bord) l'égalité (1) avec $k=2$ implique déjà le difféomorphisme, tout au moins si n est assez grand [9].

Nous allons nous occuper presque exclusivement du problème « compact ».

2. L'hypothèse de Poincaré

Toute cette problèmatique « compacte » discutée plus-haut est apparue au sujet de l'hypothèse de Poincaré. D'ailleurs, le fait qu'en essayant de prouver $M_1^n = M_2^n$(2), l'égalité répondant à ce problème, on est tombé sur

(1) Ici $\neq$ signifie non homéomorphe.

(2) Dans tout ce paragraphe on considérera, sauf mention contraire que $M_1^n = S^n$, $V^n = M_2^n$ = une variété ayant le type d'homotopie de S^n. Evidemment $S^n = V^n$ signifie ici homéomorphisme.

des égalités de type (1) plus ou moins fortes (c'est-à-dire avec k plus ou moins grand), n'est pas sans relations avec l'autre fait que l'hypothèse généralisée, en dimensions suffisamment grandes s'est avérée, tout récemment, faisable. (Voir là-dessus les beaux travaux de S. Smale [21]–[24], J. Stallings [25]–[26], A. H. Wallace [39], E. C. Zeeman [34]–[35]). Il n'est pas tout à fait inexact de dire que le problème d'abaisser la limite inférieure du domaine de dimensions où l'hypothèse de Poincaré est déjà prouvée, est du même ordre de difficulté qu'abaisser k dans une égalité de type (1).

En tout cas, cette échelle qui monte d'une égalité de type (1) avec k grand jusqu'au point final $k=0$ ($S^3=V^3$), où la difficulté de passer de k à $k-1$ croît d'une façon exponentielle quand k décroît, est l'idée générale du programme de la démonstration de l'hypothèse de Poincaré que je vais exposer maintenant. Ce programme n'est qu'en partie effectivement réalisé et je m'efforce de le mener à bout.

En général, pour prouver une égalité de type (1) on a toujours utilisé une construction du type suivant(1) : On impose tout d'abord à M_1^n, M_2^n la condition d'avoir le même type simple d'homotopie (voir [30–32]), ce qui est réalisé par exemple s'ils ont le même type d'homotopie et si $\pi_1=0$, Z_2, Z_3, Z_4, Z.

Alors il existe une chaîne finie de complexes reliant M_1^n, M_2^n : $M^n=M^1$, $M^2, \ldots M^{(f)}=M_2^n$ (qui ne sont plus des variétés, sauf aux cas extrêmes) qui se déduisent l'un de l'autre par des opérations élémentaires simples. Pour émousser les singularités de $M^{(i)}$ on doit le grossir un peu, afin de la transformer en une variété $\theta(M^{(i)})$, qui apparait comme une sorte de « voisinage » (neighbourhood) de $M^{(i)}$. Le facteur $M^{(i)} \to \theta(M^{(i)})$ doit satisfaire à $\theta(M^{(i)}) = \theta(M^{(i+k)})$ et à $\theta(M_j^n)=M^n \times D_j^k$. En générale pour que cette dernière condition puisse être satisfaite on doit imposer à M_1^n, M_2^n une condition de compatibilité relative aux fibrés tangents, qui est automatiquement satisfaite pour $n=3$. Bien entendu dans les considérations faites ci-dessus M_1^n, M_2^n n'ont pas obligatoirement le type d'homotopie de S^n.

Maintenant, pour réaliser cette idée on doit se fixer sur le choix des opérations élémentaires et du foncteur θ. Ceci est un problème technique trop compliqué pour que je puisse entrer dans des détails ici. La technique initialement utilisée par Whitehead donnait pour k des valeurs trop grandes.

Dans [17], j'ai prouvé qu'il existe un nombre $n(V^3)$ tel que

$$(S^3-\Sigma)\times D^2=(V^3-\Sigma)\times D^2, \tag{3}$$

où Σ est la réunion des intérieurs de $n(V^3)$ disques 3-dimensionnels, disjoints deux à deux, plongés d'une façon C^∞. (3) représente une approximation à un facteur [0, 1] près de l'hypothèse de Poincaré qui est équivalente à

$$(S^3-\Sigma)\times D^1=(V^3-\Sigma)\times D^1. \tag{3'}$$

On peut se faire une idée de la distance qui sépare (3) de (3′) en invoquant

(1) Pour la première fois ce type de construction apparaît chez J. H. C. Whitehead [30]–[32]. Beaucoup plus tard, avec de notables changements B. Mazur [8]–[10] et moi-même [17] avons utilisé des procédés similaires ; nos deux recherches ont été indépendantes mais il s'est trouvé qu'elles avaient beaucoup de parenté. D'autre part, quand j'ai écrit [17], je ne savais encore rien de la théorie de Whitehead, et c'est Mazur qui m'a montré tout le profit que j'en pouvais tirer.

le fait qu'il existe des variétés compactes $V^4 \neq D^4$(1) telles que $V^4 \times D^1 = D^5$ (voir [16, 19, 7])(2).

Pour prouver (3) j'ai utilisé une certaine catégorie $\mathcal{C}$ de complexes à 3 dimensions, avec singularités d'un certain type et un foncteur $K \to \theta(K)$ sur cette catégorie, à valeurs dans la catégorie des variétés à 5 dimensions. Le foncteur θ a toutes les propriétés algebriques-fonctionnelles du foncteur $K \to K \times D^2$ et coincide avec ce dernier exactement pour les variétés (donc pour la partie non singulière de $\mathcal{C}$).(3)

En dehors de ce foncteur, [17] utilise des techniques du type théorie des immersions, que j'ai reprises dans [18].

En améliorant les techniques de [17] on peut montrer que $n(V^3) = 1$, c'est-à-dire que

$$(V^3 - \operatorname{int} D^3) \times D^2 = D^5 \tag{4}$$

Les améliorations dont je parle consistent principalement dans la considération de complexes K, donnés d'une relation $R \subset K \times K$ symétrique pour laquelle deux points $x, y \in K$, tels que $(x,y) \in R$ seront considérés dans un certain sens « infiniment voisins ». Une théorie des « voisinages » $\theta(K)$ de tels complexes est developpée. De même une théorie de l'homotopie simple (de « déformations élémentaires » de Whitehead).

L'étape suivante est de prouver que $n(V^3) = 0$, donc que

$$S^3 \times D^2 = V^3 \times D^2 \tag{5}$$

qui implique

$$S^3 \times D^1 = V \times S^1 \tag{5'}$$

Pour arriver à ce but, on doit prouver que le plongement :

$$S^2 = \partial(V^3 - \operatorname{int} D^3) \xrightarrow{i} \partial((V^3 - \operatorname{int} D^3) \times D^2) = S^4 \tag{6}$$

donné par $i(p) = p \times q$ ($q \in D^2$ fixé) est trivial. Bien entendu on aimerait prouver ceci comme cas particulier de la conjecture suivante :

Si $i : S^2 \to S^4$ est un plongement différentiel tel que $\pi_1(S^4 - i(S^2)) = Z$, i est trivial(4).

En fait au lieu de cette conjecture difficile il suffirait de prouver l'une des deux conjecture suivantes, qui semblent faisables :

Soit W^3 une variété à trois dimensions et i_0, i_1 deux plongements $S^2 \to W^3$, régulièrement homotopes (en tant qu'immersions) [20] tels que i_1 est trivial. Alors le plongement $i_0 \times \frac{1}{2} : S^2 \to W^3 \times [0,1]$ est trivial.

Soit W'^3 une variété à trois dimensions telle que $\partial W'^3 = S^2$. Considérons l'espace quotient $S^1 \times W'^3/\phi$, ou (ϕ) est la relation suivante :

si $p, q \in S^1$, $x \in S^2 = \partial W'^3$ $(p,x) \equiv (q,x)$

(1) $V^4 \neq D^4$ signifie non homéomorphe.

(2) De telles décompositions non triviales pour $D^n (n > 5)$ ont été construites par M. Curtis [3], qui utilise un espace de M. H. Newman [15]. R. H. Bing [2] a prouvé que pour $n = 4$ n'existent que des décompositions triviales.

(3) J'ai appelé θ « le produit cartésien régularisé par D^2 ».

(4) On a des résultats similaires, dus à J. Stallings [26] (voir aussi [36]), mais en dimensions plus élevées et la trivialité est topologique, non différentielle.

La proposition énoncée semble très difficile. En tout cas elle sort du domaine de dimensions ou la théorie de Haefliger [5] est appliquable.

On a de bonnes raisons pour croire qu'un tel résultat n'est pas vrai, sous forme différentielle, en dimensions élevées.

Pour la dimension $n = 3$ c'est le « Hauptsatz der Knotentheorie », prouvé dans un travail classique par C. D. Papakyriakopoulos.

Soit $\pi: S^1 \times W'^3 \to S^1 \times W'^3/(\phi)$ l'application canonique. Supposons que $S^1 \times W'^3/(\phi) = S^4$, alors le plongement $\pi: p \times S^2 (= \partial W'^3) \to S^4$ est trivial.

Mais je pense prouver que le plongement i de (6) est trivial de la façon suivante. Tout d'abord au lieu de V^3 je considererai $V^3 + (-V^3) = {}^*V^3$ ($+$ veut dire somme connexe). Comme l'a remarqué J. Milnor [11], pour prouver l'hypothèse de Poincaré il suffit de montrer que ${}^*V^3 = S^3$. Je désignerai par ${}^*S^2 \subset {}^*V^3$ la sphère S^2 vis-à-vis de laquelle ${}^*V^3$ est naturellement symétrique. $[(V^3 - \operatorname{int} D^3) + (-(V^3 - \operatorname{int} D^3))] \times D^2 = V'$ peut être factorisé comme $S^2 \times D^3$ de deux façons différentes : ou bien en lui appliquant directement les techniques qui ont permis de prouver (4) (et alors on obtient une factorisation en un certain sens symétrique par rapport à S^2) ou bien en retranchant de $D^5 = ({}^*V^3 - \operatorname{int} D^3) \times D^2 = [(V^3 - \operatorname{int} D^3) + (-V^3)] \times D^2$, $\operatorname{int} D^3 \times D^2 \subset (-V^3) \times D^2$. La trivialité de i revient principalement, à ce que ces deux factorisations coincident en un certain sens naturel à expliciter. Or, ces deux factorisations ont été obtenues par des suites d'opérations élémentaires qui « peuvent être comparées pas à pas ». Cette étape semble donc faisable.

L'égalité (5′) une fois établie on y déduit immédiatement que la souspension sur ${}^*V^3$ est S^4. En tenant compte de tout ce qui est différentiable dans cette identification, ${}^*V^3$ apparaît comme un niveau quelconque d'une fonction Φ définie sur E^4, qui satisfait aux conditions suivantes : — Φ est continue et de classe C^∞ sur

$$E^4 - 0$$

$$-\Phi(0) = 0$$

$$-\Phi \text{ a un minimum en } 0$$

$$-\operatorname{grad}\Phi \neq 0 \text{ sur } E^4 - 0.$$

Φ est bien entendu le paramètre de la souspension. Maintenant, tout est fini, si Φ peut être régularisée, dans le sens suivant : trouver une fonction Ψ qui ne diffère de Φ que dans un voisinage de 0, qui soit de classe C^3, avec un minimum non degéneré en 0, et telle que grad $\Psi \neq 0$ sur $E - 0$. Ce problème est en général une inégalité différentielle globale fort difficile [29]. Mais on peut toujours la résoudre, sous certaines conditions suffisantes, du type suivant :

$$-0 < m \leqslant \|\operatorname{grad}\Phi\| \leqslant M < \infty$$

$$-\left|\frac{\partial^2\Phi}{\partial x^i \partial x^j}\right|, \quad \left|\frac{\partial^3\Phi}{\partial x^i \partial x^j \partial x^k}\right| \leqslant M$$

(tout ça au voisinage de 0). Il suffirait de montrer qu'on peut choisir Φ, de telle sorte que ces conditions soient remplies.

Φ est complètement determiné par un difféomorphisme ϱ du revêtement universel de $S^1 \times {}^*V^3, X^4$ sur $E^4 - 0$ et par une fonction χ définie sur X^4, de classe C^∞, avec grad $\chi \neq 0$ et qui admet comme niveaux les diverses variétés à trois dimensions obtenues en relevant $x \times {}^*V^3$ ($x \in S^1$) dans X^4. On peut choisir ϱ tel que pour tout point $p \in X^4$ la transformation infinitésimale $d\varrho_p$ admet une borne supérieure uniforme pour les rapports des modules de ses valeurs propres. Ceci fait que si l'on choisit χ de telle manière que la condition vis-à-vis des dérivées de premier ordre soit satisfaite elle le sera aussi pour Φ.

On remarque ensuite que X^4 a une métrique riemannienne naturelle pour laquelle les niveaux de χ ont les courbures uniformément bornées. Ceci permet de faire en telle sorte que la condition vis-à-vis des derivées de second ordre soit aussi satisfaite. On pourrait faire des considérations analogues pour les

$$\frac{\partial^3 \Phi}{\partial x^i \partial x^j \partial x^k}.$$

Cette dernière étape paraît donc aussi faisable.

Il est facile de voir que si ce programme sera complètement réalisé, il donnera aussi la réponse en dimension 4.

Remarque 1. Il semble très plausible que les méthodes qui ont servi à prouver (4) permettent de prouver que si V^5 est obtenu en grossissant un complexe contractible (compact) à deux dimensions $V^5 = D^5$. D'ici on pourrait résoudre un problème de de Rham en prouvant que $V_1 \times I_1 = D^5$ (voir [19], p. 7). D'ailleurs, $V_1 \times I_1 \neq D^5$ entraînerait des conséquences bien fâcheuses.

Remarque 2. Je voudrais insérer ici quelques remarques sur les variétés à 3 dimensions qui n'ont pas de relations directes avec les considérations ci-dessus.

On peut définir d'une façon naturelle un nombre $N(W^3)$ attaché à une variété à 3 dimensions quelconque : le « degré de complication de W^3 ». Par exemple on peut procéder ainsi : on part d'une triangulation(1) donnée de W^3. Soit C une courbe simple, homotope à 0, de la triangulation. Considérons un disque singulier de frontière C. Ce disque peut être toujours pris assez régulier pour qu'il se contracte soit à un point soit à un bouquet de cercles. S'il se contracte à un point, tout est fini, sinon on considère des disques singuliers ayant comme frontières ceux des cercles du bouquet qui sont homotopes à 0. Ces disques peuvent être pris tels que la figure se contracte à un point ou à un bouquet de cercles. On continue jusqu'à ce qu'on arrive soit à un point soit à un bouquet de cercles tous non homotopes à 0. Le nombre minimum d'étapes nécessaires pour arriver à ce but sera désigné par $N(C)$.

Par définition:

$$N(W_3) = \inf \sup \{N(C)\},$$

où $N(C)$ sont tous les $N(C)$ des C d'une certaine triangulation et l'inf est pris par rapport à toutes les triangulations. Il semble bien que pour des complémentaires d'arcs de Fox et Artin le $N = \infty$ Mais on peut faire la conjecture suivante, beaucoup plus générale que l'hypothèse de Poincaré qui en est un cas très particulier :

Si W^3 est compact $N(W^3) < \infty$.(2) Cette conjecture prouvée, on pourrait y déduire, en dehors de l'hypothèse de Poincaré la conjecture suivante :

Soit $S^1 \subset S^3$ un noeud. Le revêtement universel de $S^3 - S^1$ est homéomorphe à E^3.

---

(1) On admet des simplexes à 3 dimensions dont une face manque.

(2) Donc : « Une variété compacte ne peut être infiniment compliquée », ce qui semble raisonnable.

3. La théorie de B. Mazur

Je vais énoncer dans ce qui s'ensuit quelques-uns des principaux résultats de B. Mazur concernant le problème « compact » de l'introduction ([8]–[10]).

Soit X un complexe. Un « voisinage » (neighbourhood) de X sera une variété compacte, obtenue en grossissant X. En réalité le traitement « Morse Theoretic » de Mazur est beaucoup plus élégant : à la variété différentielle il substitue une «filtration différentielle » (differential filtration) qui est quelque chose comme la décomposition d'une variété differentielle par une fonction non dégénerée. Au complexe il substitue une « filtration cellulaire » (cell filtration) qui est quelque chose comme une décomposition de Thom attachée à une filtration différentielle du type défini plus haut. Mazur dit qu'une filtration différentielle « voisinage » d'une filtration cellulaire X doit être conçue comme un relèvement de X dans le royaume de la topologie différentielle, comme une généralisation d'un fibré à fibre vectorielle sur X, comme une stabilisation de X, et aussi comme un voisinage, au sens ordinaire du mot, de X, plongé dans quelque chose.

On verra que la théorie de Mazur établit un parallélisme entre problèmes « compacts » et homotopie simple, entre problèmes « ouverts » et homotopie habituelle.

Soit $N^n(X)$ l'ensemble des voisinages n-dimensionnels de X. On montre que n suffisamment grand $N^n(X)$ est en correspondance biunivoque avec le groupe de Grothendieck des fibrés à fibre vectorielle sur X. $N^n(X)$ est fonctorielle par rapport à l'homotopie simple. C'est-à-dire que toute homotopie simple $f:X\to Y$ induit une bijection $f_*:N^n(X)\to N^n(Y)$ et $N\in N^n(X)$ et f_*N sont difféomorphes. Ceci, dans le langage de Mazur est un résultat « stable ». On obtient un résultat « non stable » correspondant (The nonstable neighbourhood theorem) en indiquant la borne inférieure des n. (max $(\dim X, \dim Y)+2$).

Introduisons maintenant la définition suivante :

Deux variétés M_1^n, N_2^n seront dites avoir le même type d'homotopie (simple) de Mazur, s'il existe une équivalence homotopique (simple) $f:M_1^n \to M_2^n$ telle que si $T(M_1^n), T(M_2^n)$ sont les fibrés tangents respectifs:

$$f^*T(M_2^n)\oplus O = T(M_1^n)\oplus O,$$

ou $\oplus$ est la somme de Whitney et O un fibré trivial. Alors les considérations plus haut citées, permettent de prouver le théorème suivant :

Soient M_1^n, M_2^n deux variétés ayant le même type simple d'homotopie de Mazur. Alors, pour k suffisamment grand, on a (1). ($k>n$; mais il semble que par la même méthode cette dimension puisse être beaucoup plus abaissée; probablement jusq'à 3.)

En tout cas, la condition de compatibilité relative aux espaces tangents est absolument essentielle ici. Autrement on en déduirait que les classes de Pontrjagin sont des invariants du type simple d'homotopie, ce qui est faux, comme on le voit par exemple en utilisant [28].

Des résultats analogues au précédent se trouvent chez M. Hirsch [6]. Si l'on met au lieu du même type simple d'homotopie, même type d'homotopie et $\pi_1=0$, le théorème apparaît comme corollaire d'un résultat trés général de Smale [24] :

Soient $W^n \subset M^k$ avec $n > 5$, $\partial M = 0$, $\partial W \neq 0$, $\pi_2(\partial W) = \pi_1(M) = 0$, l'inclusion étant équivalence homotopique. Alors W est un fibré à disques sur M.

Hirsch prouve aussi que si M_1^n, a le type d'homotopie de S^n, et est une π-variété(1) on a (1) avec $k > \frac{1}{2}(n+1)$. ($M_2^n = S^n$); de plus M_1^n est frontière d'une variété parallélisable en peut mettre $k = 3$.

Revenons à la théorie de Mazur. Le « non-stable neighbourhood theorem » est un instrument puissant à l'aide duquel on peut prouver par exemple des résultats du type suivant : Soit M^n une variété sans bord, $n \geqslant 6$. Supposons qu'il existe un complexe X qui n'a qu'une cellule de dimension n, aucune de dimension $n-1$, $n-2$, ayant le même type d'homotopie que M. M est alors un voisinage de X. Ceci implique l'hypothèse de Poincaré pour $n \geqslant 6$.

Mazur introduit aussi le concept de « voisinage simple » de X. Soit $f: X \subset M^n$ un plongement régulier. Une sousvariété compacte $U^n \subset M^n$ telle que, $f: X \subset U^n$ est une équivalence homotopique simple, $f(X) \subset$ int U^n, $\pi_i(U - f(X), \partial U) = 0$ $(i = 1, 2)$, sera appellée un voisinage simple de x.

Si $n \geqslant 6$, U est complètement determiné par f. On en déduit que deux variétés s-cobordantes de dimension $\geqslant 6$ sont difféomorphes, ce qui généralise un théorème de Smale [24]. (Le s-cobordisme est ce que l'on obtient du h-cobordisme(2) en substituant l'homotopie simple à l'homotopie habituelle).

4. Les cas ouvert

Si l'on remplace dans la théorie de Mazur l'équivalence homotopique simple par l'équivalence homotopique habituelle on tombe sur le théorème suivant :

Si M_1^n, M_2^n sont deux variétés ayant le même type d'homotopie de Mazur, (2) est vrai pour k assez grand. Un ces particulier de ce théorème de Mazur est prouvé dans [12]. La démonstration pour le cas général est très analogue. Elle fait usage tout à fait comme dans [12] d'un raisonnement de type « infini ».

Je vais trés brièvement rappeller quelques autres résultats « ouverts ». On a tout d'abord le théorème de Stallings qui affirme que si V_1^n, V_2^n sont h-cobordantes et $n \geqslant 4, V_1^n \times E^1 = V_2^n \times E^1$. Ce théorème est combinatoire et non différentiel. Bien entendu il est interessant si le h-cobordisme n'est pas s-cobordisme.

C'est toujours Stallings qui a prouvé que si M^n est une variété ouverte, contractible, simplement connexe à l'infini, $n \geqslant 5, M^n = E^n$. Ceci implique que si M^n, M^m sont ouvertes et contractibles, $n, m \geqslant 1$, $n + m \geqslant 5$, alors $M^n \times M^m = E^{n+m}$ [27](3). Ceci généralise un résultat antérieur de McMillan et Zeeman [14].(4)

Je vais terminer en citant quelques exemples.

Bing a construit un espace B qui n'est pas variété, tel que $B \times E^1 = E^4$ [1]. Whitehead a introduit le classe suivante d'espaces : une variété ouverte, contractible, V^3, sera appellée W-espace si tout compact $K \subset V^3$ peut être plongé dans E^3. Cette classe contient des éléments nontriviaux [34].

Si $U \subset W$, $U \times E^1 = E^4$; Si $U_1, U_2 \in W$, $U_1 \times U_2 = E^4$ [4], [13].

(1) C'est-à-dire si le fibré normal de M_1^n plongé dans E^{n+M} est trivial (M grand).

(2) Ou J-équivalence, comme on disait il y a quelque temps.

(3) En réalité les résultats de Stallings sont combinatoires. Ils deviennent différentiels en utilisant un résultat de Thom, Milnor et Munkers.

(4) Pour des résultats topologiques seulement, voir le travail tout récent de M. W. Hirsch [38].

BIBLIOGRAPHIE

[1]. BING, R. H., The cartesian product of a certain nonmanifold and a line is E^4. *Ann. Math.*, 70 (1959), 399–452.

[2]. BING, R. H., A set is a 3-cell if its cartesian product with an arc is a 4-cell. *Proc. Amer. Math. Soc.*, 12 (1961), 13–19.

[3]. CURTIS, M. L., Cartesian products with intervals. *Proc. Amer. Math. Soc.*, 12 (1961), 819–820.

[4]. GLIMM, J., Two cartesian products which are euclidean spaces. *Bull. Soc. Math. France*, 88 (1960), 131–135.

[5]. HAEFLIGER, A., Differential imbeddings. *Bull Amer. Math. Soc.*, 67 (1961), 109–112.

[6]. HIRSCH, M. W., Diffeomorphisms of tubular neighbourhood. (A paraître.)

[7]. MAZUR, B., A note on some contractible 4-manifolds. *Ann. Math.*, 73 (1961), 221–228.

[8]. MAZUR, B., Stable equivalence in differential topology. *Bull. Amer. Math. Soc.*, 67 (1961), 377–384.

[9]. MAZUR, B., Simple neighbourhoods. *Bull. Amer. Math. Soc.*, 68 (1962), 87–92.

[10]. MAZUR, B., The theory of neighbourhoods. (A paraître.)

[11]. MILNOR, J., *Differential Manifolds which are Homotopy Spheres*. Princeton, 1959 (Mimeographic.)

[12]. MILNOR, J., Two complexes which are homeomorphic but combinatorially distinct. *Amer. J. Math.*, 74 (1961), 575–590.

[13]. MCMILLAN, D. R., Cartesian products of open manifolds. *Bull. Amer. Math. Soc.*, 67 (1961), 510–519.

[14]. MCMILLAN, D. R. & ZEEMAN, E. C., On contractible open manifolds. (A paraître.)

[15]. NEWMAN, M. H. A., Boundaries of ULC sets in euclidean n-space. *Proc. Nat. Acad. Sci.* 34 (1948), 193–196.

[16]. POÉNARU, V., Les décompositions de l'hypercube en produit topologique. *Bull. Soc. Math. France*, 88 (1960), 113–129.

[17]. —— Sur quelques propriétés, des variétés simplement connexes à trois dimensions. *Rend. di Mat.*, 20 (1961), 235–269.

[18]. —— Sur la théorie des immersions. (A paraître.)

[19]. DE RHAM, G., Factorisations topologiques du disque à cinq dimensions. *Séminaire Ch. Ehresmann*, Paris 1961.

[20]. SMALE, S., The classification of immersions of spheres in euclidean spaces. *Ann. Math.*, 69 (1959), 327–344.

[21]. SMALE, S., The generalized Poincaré conjecture in higher dimensions. *Bull. Amer. Math. Soc.*, 66 (1960), 373–375.

[22]. —— Generalized Poincaré conjecture in dimensions greater than four. *Ann. Math.*, 74 (1961), 391–406.

[23]. —— Differential and combinatorial structures on manifolds., *Ann. Math.*, 74 (1961), 498–502.

[24]. —— On the structure of manifolds. (A paraître.)

[25]. STALLINGS, J., Polyhedral homotopy-spheres. *Bull. Amer. Math. Soc.*, 66 (1960), 485–488.

[26]. —— The topology of high-dimensional piece-wise linear manifolds. (A paraître.)

[27]. —— The piece-wise linear structure of euclidean space. (A paraître.)

[28]. TAMURA, I., On Pontrjaghin classes and homotopy types of manifolds. *J. Math. Soc. Japan*, 9 (1957), 250–262.

[29]. THOM, R., Remarques sur les problèmes comportant des inéquations différentielles globales. *Bull Soc. Math. France*, 87 (1959), 455–461.

[30]. WHITEHEAD, J. H. C., Simpliciel spaces, Nuclei, and m-groups. *Proc. London Math. Soc.*, 45 (1939), 243–327.

[31]. —— Simple homotopy types. *Amer. J. Math.*, 72 (1950), 1–57.

[32]. —— *Simple Homotopy Types*. Oxford, 1959. (Mimeographic.)

[33]. —— A certain open manifold whose group is unity. *Quart. J. Math.*, 6 (1935), 268–279.

[34]. ZEEMAN, E. C., The generalized Poincaré conjecture. *Bull. Amer. Math. Soc.*, 67 (1961), 270.

[35]. —— The Poincaré conjecture for $n \geqslant 5$. dans le volume: *Topology of 3-manifolds and related Topics* (édité par M. K. Fort Jr.). Georgia, 1962, 198–204.

[36]. —— Knotting manifolds. *Bull. Amer. Math. Soc.*, 67, (1961), 117–119.

[37]. ZEEMAN, E. C., Voir McMillan.

(Ajouté aux épreuves) :

[38]. HIRSCH, M. W., On the product of a contractible topological manifold and a cell. (A paraître.)

[39]. WALLACE, A. H., Modifications and cobounding manifolds. *J. Math. Mech.* 10 (1961), 773–809.

DYNAMICAL SYSTEMS AND THE TOPOLOGICAL CONJUGACY PROBLEM FOR DIFFEOMORPHISMS

By S. SMALE

For simplicity we consider an ordinary *differential equation* (or a dynamical system) to be a C^r vector field on a C^r manifold which generates a 1-parameter group of diffeomorphisms. An *equivalence* (or topological equivalence) between two differential equations is a homeomorphism preserving sensed trajectories (or orbits). The qualitative problem of differential equations is to obtain information on equivalence classes of differential equations on a given manifold.

This motivates us to consider the topological conjugacy problem for diffeomorphisms (a diffeomorphism is a differentiable homeomorphism with a differentiable inverse). More precisely, we say two diffeomorphisms $T_1, T_2: M \to M$ (say C^r diffeomorphisms of a C^r manifold, r always positive) are topologically conjugate if there exists a homeomorphism $h: M \to M$ such that $hT_2 = T_1 h$. The problem then is to study the topological conjugacy classes of diffeomorphisms of a given manifold.

There are several reasons for studying the latter problem, the most important being the following. It appears that usually a qualitative problem in differential equations has an analogue in the conjugacy problem. This analogue is a little simpler than the original, and if solved, its solution seems to give a way of doing the original problem. In any case, everything said in what follows on the conjugacy problem can be translated into statements about differential equations. At the end of our survey we indicate how this can be done.

It should also be noted that the above problems may be viewed as special cases in the study of a non-compact Lie Group G acting differentiably on a manifold, corresponding to $G = R$ and $G = Z$.

As enunciated in [7] for differential equations, the main conjugacy problem as we see it is the following. Given compact M, let $\mathcal{D}_M$ be the space of C^r diffeomorphisms of M in the C^r topology (diffeomorphisms are C^r close if they are pointwise close together with their first r derivatives). Then one seeks an open dense subset $\mathcal{C}$ of $\mathcal{D}_M$, somehow amenable to classification (say by numerical and algebraic invariants).

A fruitful notion relative to this problem is that of a structurally stable diffeomorphism (the analogous definition for a differential equation was given by Andronov and Pontrjagin in 1937; see [3]). The h in the definition of topologically conjugate is called an equivalence between T_1 and T_2, and if h is pointwise within ε of the identity (in some fixed metric on M) it is called an ε-equivalence. Then a diffeomorphism $T_1: M \to M$, compact M, is *structurally stable* if given $\varepsilon > 0$, there exists $\delta > 0$ such that if $d_{c^1}(T_1, T_2) < \delta$ for some diffeomorphism $T_2: M \to M$, then T_1 and T_2 are ε-equivalent. Here d_{c^1} is a C^1 metric on $\mathcal{D}_M$.

The problem of structural stability (for diffeomorphisms) is: given M, are

the structurally stable diffeomorphisms dense in $\mathcal{D}_M$? If dim $M>1$, this is an open and difficult problem. In fact it is not known if there exists even one structurally stable diffeomorphism on a given manifold.

In any case it is clear that the periodic points (i.e. points $x \in M$ such that $T^m x = x$, $m \neq 0$, T^m the composition $T: M \to M$ with itself m times) and associated global stable and unstable manifolds will play a basic role in the topological conjugacy problem. So at this point we give the "stable manifold theorem". For more details and history, see [10].

Let $T: M \to M$ be a diffeomorphism with fixed point $p \in M$. The derivative of T at p is a linear automorphism of the tangent space M_p of M at p. The point p will be called an *elementary* fixed point of T if this derivative has no eigenvalue of absolute value one.

(A) STABLE MANIFOLD THEOREM. *Let p be an elementary fixed point of a C^∞ diffeomorphism $T: M \to M$ and E_1 the (eigen) subspace of M_p corresponding to the eigenvalues of the derivative of T at p of absolute value less than one. Then there is a C^∞ map $R: E_1 \to M$ which is an immersion (i.e. with Jacobian of rank* $=$ dim E_1 *everywhere*), 1–1, *and has the property $TR = RT_1$ for some contraction $T_1: E_1 \to E_1$. Also $R(p) = p$ and the derivative of R at p is the inclusion of E_1 into M_p.*

By a *contraction* we mean a diffeomorphism T_1 of E_1 onto itself such that there is a differentiably imbedded disk D in E_1 with $T_1 D \subset$ interior D, $\bigcap_{i>0} T_1^i\, D =$ origin of E_1, $\bigcup_{i<0} T_1^i\, D = E_1$. The map $R: E_1 \to M$ or sometimes the image of R is called the *stable manifold* of p, or of T at p. The *unstable manifold* of T at p is the stable manifold of T^{-1} at p.

A point $p \in M$ will be called an *elementary* periodic point of $T: M \to M$ if p is an elementary fixed point of T^m for some m. An (*elementary*) *periodic orbit* is the finite set $\bigcup_{i \in Z} T^i\, p$ where p is an (elementary) periodic point. If $T^m p = p$, m the minimal positive integer with this property, then m is called the least period of p. The definition of the stable manifold of an elementary periodic orbit is as follows. Let $\varphi: E_1 \to M$ be the stable manifold of T^m at p where m is the least period of p, p in our periodic orbit. Then $R: E_1^i \to M$ is defined by $R = T^i \varphi$ where $0 \leqslant i < m$ and E_1^i is a copy of E_1. Thus the stable manifold of a periodic orbit is a 1–1 immersion of the disjoint union of m copies of a Euclidean space. The unstable manifold of a periodic orbit is the stable manifold of the periodic orbit relative to T^{-1}.

To give some direction to our survey we list some axioms a diffeomorphism of a manifold might happen to satisfy. Fixing a compact manifold M, the space of C^r diffeomorphisms of M with the C^r topology is denoted as usual by $\mathcal{D}_M$. We will define T to be in a subspace $\mathcal{C}_M$ of $\mathcal{D}_M$ if and only if T has the following properties, i.e. T satisfies Axioms 1, 2′, 3.

Axiom 1. Every periodic point of T is elementary.

Axiom 1 implies that T has only countably many periodic points.

Axiom 2. T has the normal intersection property.

This means the following: if β_1, β_2 are elementary periodic orbits of T, let W_1^u, W_2^s be the unstable manifold of β_1, stable manifold of β_2 respectively. Then if $x \in W_1^u \cap W_2^s$ the tangent spaces of W_2^u and W_1^u at x span the tangent space of M at x.

The following is a strengthening of Axiom 2.

Axiom 2′. Let β_1, β_2, W_1^u, W_2^s be as above. Then there exists a neighborhood $\mathcal{U}$ of β_1 with the following property. Each component of $\mathcal{U} \cap W_2^s$ is a cell which has a non-empty transversal intersection with W_1^u. The same is true with W_1^u, W_2^s replaced by W_1^s, W_2^u respectively.

Axiom 3. (*a*) Let Ω be the closure of the set of periodic points of T in M. Then for every $x \in M$, limit $m \to \pm\infty$ $T_x^m \subset \Omega$.

(*b*) The union of the stable manifolds of all the periodic orbits of T is a dense subset of M. The same is true of the unstable manifolds.

We pose two questions: (*a*) Is $\mathcal{C}_M$ open, dense, in $\mathcal{D}_M$? (*b*) Is $T \in \mathcal{C}_M$ a necessary and sufficient condition that T be structurally stable?

Although very possibly, in the final picture, $\mathcal{C}_M$ will not be the structurally stable diffeomorphisms, it seems that to date it is the best guess for such ((compare [7] or [8]!).

One can study these problems from the following point of view:

I. *The approximation problem.* Approximate a given diffeomorphism by a diffeomorphism in $\mathcal{C}_M$, and

II. *The regularity problem.* Find regularity properties of elements of $\mathcal{C}_M$.

We first discuss I. One can approximate an arbitrary $T \in \mathcal{D}_M$ by T satisfying Axiom 1. In fact,

(B) THEOREM. *Let $\mathcal{E}$ be the subspace of $\mathcal{D}_M$ consisting of T with every periodic point elementary. Then $\mathcal{E}$ is the countable intersection of open and dense sets in $\mathcal{D}_M$.*

For a proof see [10]. Independently, R. Abraham has shown that this follows from a general transversality theorem [1]. The paper of L. Markus [4] is also in the direction of Theorem C.

A similar situation holds for Axiom 2.

(C) THEOREM. *Let $\mathcal{J}$ be the subspace of $\mathcal{E}$ (of the previous theorem) of diffeomorphisms with the normal intersection property. Then $\mathcal{J}$ is the countable intersection of open and dense sets in $\mathcal{D}_M$.*

See [10] for a proof.

Unfortunately, there is no similar theorem known for Axiom 2′.

In Axiom 3, parts (*a*) and (*b*) seem to be related in some fashion, but it is not clear how. Does (*a*) imply (*b*), or conversely? To approximate a given diffeomorphism by one satisfying 3*a* or 3*b* is a central problem, related to what is sometimes called the problem of the "closing Lemma". See Peixoto [5] for an account of this important problem.

A special case of our approximation problem is the following.

(D) *Problem.* Let $T: M \to M$ be a diffeomorphism of a compact manifold. Is there a C' approximation T' of T such that T' has a periodic point?

The answer is not known for the 2-dimensional torus.

In the discussion of the regularity problem stated above, we start with the case of diffeomorphisms satisfying a highly restrictive axiom in addition to Axioms 1, 2′, 3.

Axiom 4. T has a finite number of periodic points.

The set of diffeomorphisms satisfying Axioms 1, 2′, 3, 4 is denoted by $\mathcal{C}_M^4$.

It is important to note that in assuming Axiom 4, an open set of diffeomorphisms of $\mathcal{D}_M$ is lost (see e.g. [11] or below). On the other hand one can say some substantial things about elements of C_M^4.

(E) THEOREM. *Diffeomorphisms in C_M^4 satisfy a form of the Morse inequalities relating the periodic points. The union of the stable manifolds associated to the periodic points is all of M. The boundary of a stable manifold of dimension p is the union of stable manifolds of dimension $\leqslant p$.*

The proof is essentially contained in [8]. A task which seems important and yet tractable is to show that the $T \in C_M^4$ are structurally stable. However this has not even been carried out under the additional assumption of Axiom 5.

Axiom 5. Let W_i^u, W_j^s be an unstable, stable manifold respectively of periodic orbits of T which have non-empty intersection. Then

$$\dim W_i^u + \dim W_j^s > \dim M.$$

We say that the set of T satisfying all of our axioms is C_M^5. The following can be proved along the lines of [8], [9].

(F) THEOREM. *On every manifold there exist non-empty open sets of $\mathcal{D}_M$ which are contained in C_M^5. If $T \in C_M^5$, the boundary of a stable manifold of T is the union of lower dimensional stable manifolds. The components of the stable manifolds generate the homology of M in a natural way and the correspondence between the stable manifold of a periodic point and the unstable manifold induces Poincaré duality. The Morse inequalities in the previous theorem can be interpreted to include the usual ones.*

This theorem shows that C_M^5, and hence C_M^4, C_M are not empty for any M. Thus if it could be shown that every $T \in C_M^5$ is structurally stable, we would have proved that there exist structurally stable diffeomorphisms on every compact manifold.

Next the question comes up as to the existence of elements of C_M which are not in C_M^4, i.e. those with an infinite number of periodic points. The following is in [11].

(G) THEOREM. *There exist open sets in $\mathcal{D}_M$ for M an aribitrary n-sphere, $n > 1$, with the following properties:* (1) *the diffeomorphisms are in C_M;* (2) *they are structurally stable;* (3) *the diffeomorphisms have an infinite number of periodic points* (*and minimal sets homeomorphic to a Cantor set*).

This theorem answers the question as to whether a structurally stable diffeomorphism (differential equation) can have an infinite number of periodic points (closed orbits). It seems that this theorem and its proof are quite important. It shows that one can cope successfully with difficult phenomena present in differential equations of dimension greater than two and not present in two-dimensional differential equations.

In the examples of Theorem G, one has present homoclinic points. A *homoclinic point* associated to a periodic orbit β of a diffeomorphism is a point of intersection of the stable and unstable manifolds associated to β. Homoclinic points were first discovered by Poincaré [6] in the restricted 3-body problem, and studied by Birkhoff [2]. Merely the existence of a homoclinic point implies considerable complications. At the end of his

three volumes on celestial mechanics [6, p. 389], referring to homoclinic points, Poincaré wrote: "On sera frappé de la complexité de cette figure, que je ne cherche même pas à tracer. Rien n'est plus propre à nous donner une idée de la complication du problème des trois corps et en général de tous les problèmes de Dynamique ....."

The methods used in proving Theorem G not only completely describe the homoclinic situation in those examples, but can be applied to give some understanding of arbitrary homoclinic points as well.

A different example which exhibits stable manifolds and homoclinic points has a simple description. In the plane E^2 let T_0 be a linear automorphism given by a 2×2 matrix with integer entries, determinant ± 1, and an eigenvalue greater than one in absolute value. Then T_0 induces a diffeomorphism T of the torus. The origin of E^2 projects into a fixed point p of T and the eigendirections of T_0 project into the stable and unstable manifolds of p. One easily sees that the homoclinic points associated to p are dense in the torus.

(H) THEOREM. *The T described above is a structurally stable diffeomorphism of the torus.*

After a meeting in September 1961 in Kiev on non-linear oscillations (where I announced Theorem G in dimension 2 [14]). I visited Moscow and spoke with mathematicians D. V. Anosov, V. I. Arnold, and Y. G. Sinai, among others. There I conjectured Theorem H and that geodesic flows on compact Riemannian manifolds of negative curvature were also structurally stable. Since then I and (as I have learned at this Congress), Arnold and Sinai have independently proved Theorem H, the latter proof having the advantage of being published [13]. In addition, Anosov has very recently proved (as I have learned also at this Congress) a beautiful theorem which settles the above conjecture for geodesic flows affirmatively and gives the n-dimensional generalization of Theorem H [12].

It seems likely that if $T \in \mathcal{C}_M$ and $T \in \mathcal{C}_M^4$, then T has homoclinic points. In the same vein one can ask if Axiom 5 implies Axiom 4 (certainly the converse is false).

An elementary periodic point will be called *elliptic* if all the associated eigenvalues have absolute value less than one, or all greater. It seems resonable to expect that if $T \in \mathcal{C}_M$, then T will have only a finite number of elliptic points. It would be nice to have a proof of this. Also if $T \in \mathcal{C}_M$ where M is the 2-sphere, must T have at least one elliptic point?

We indicate briefly how the previous discussion goes over into the analogous situation for differential equations. The definition of a structurally stable differential equation is exactly the same as for diffeomorphisms. The problem of structural stability for differential equations asks if the structurally stable differential equations on a compact manifold M are dense in the Banach space $\mathcal{B}_M$, C^r norm, of all C^r differential equations on M. This has been answered in the affirmative if dim $M \leqslant 2$ by M. Peixoto, see Theorem I, below. One constructs stable and unstable manifolds for differential equations associated to each singular point and each closed orbit of a general type (corresponding to elementary periodic points). For details see [10]. The analogues of the previous axioms can be stated for this case. The union of the singular points and closed orbits of the differential equation replace the periodic points of the diffeomorphism. Let $\mathcal{B}_M^*$ be the subspace of $\mathcal{B}_M$ of

differential equations satisfying the analogues of axioms 1, 2′, 3. The elements of $\mathcal{B}_M^*$ will have only a finite number of singular points but may have an infinite number of closed orbits if dim $M > 2$.

For the 2-dimensional case there is the following important theorem of Peixoto [5].

(I) THEOREM. *Let M^2 be a compact 2-manifold. Then (a) X on M is structurally stable if and only if $X \in \mathcal{B}_M^*$; (b) $\mathcal{B}_M^*$ is open and dense in $\mathcal{B}_M$.*

Also the analogues of the previous theorems are all valid for the differential equations case.

We comment on the problem of the existence of a first integral of a differential equation. A *first integral* of X on M in a C^r function $f: M \to R$ such that f is constant on each orbit but not on any open set. Problem E has the following analogue.

(E*) *Problem.* Can every non-singular vector field on a compact manifold be C' approximated by one with a closed orbit?

If Problem E* has an affirmative solution it follows that:

(J) The subset of $X \in \mathcal{B}_M$ with an elementary closed orbit is open and dense in $\mathcal{B}_M$.

Here *elementary* corresponds to elementary periodic point and the precise definition is in [10]. Putting J together with the analogue of Theorem B, as has been essentially observed by R. Thom, we obtain:

(J*) There exists a subset of $X \in \mathcal{B}_M$ with no first integral which is open and dense in $\mathcal{B}_M$.

A relation between the topological conjugacy problem for diffeomorphisms and the equivalence problem of differential equations is given by cross-sections, used by Poincaré and Birkhoff (e.g. see [2] or [10]). In a different direction, since a differential equation generates a 1-parameter group, one may ask under what conditions, is a diffeomorphism imbeddable in a flow? The following is not difficult.

(K) THEOREM. *Let $T \in \mathcal{C}_M$. If T can be imbeeded in a flow, then T satisfies Axioms 4, 5, every periodic point is fixed and in the neighborhood of every fixed point, T is imbeddable in a flow.*

It would be interesting to know if the converse is true.

Certainly the main problems stated here are very difficult. On the other hand, it seems quite possible to us that this field may develop rapidly and already as indicated here, there have been some initial steps in this direction.

REFERENCES

[1]. ABRAHAM, R., Transversality of manifolds of mappings. (To appear.)

[2]. BIRKHOFF, G. D., *Collected Mathematical Papers*. New York, 1950.

[3]. LEFSCHETZ, L., *Differential Equations: Geometric Theory*. Interscience Publishers, New York, 1957.

[4]. MARCUS, L., Structurally stable differential systems. *Ann. Math.*, 73 (1961), 1–19.

[5]. PEIXOTO, M., Structural stability on 2-dimensional manifolds. *Topology*, 2 (1962), 101–121.

[6]. POINCARÉ, H., *Les Méthodes Nouvelles de la Mécanique Céleste*, Vols. I–III. Gauthier-Villars, Paris, 1899. (Reprinted New York, 1957.)

[7]. SMALE, S., On dynamical systems. *Boletin de la Sociedad Matematica*, 1960, 195–198.

[8]. —— Morse inequalities for a dynamical system. *Bull. Amer. Math. Soc.*, 66 (1960), 43–49.

[9]. —— On gradient dynamical systems. *Ann. Math.*, 74 (1961), 199–206.

[10]. —— Stable manifolds for differential equations and diffeomorphisms. (To appear.)

[11]. —— Diffeomorphisms with many periodic points. (To appear.)

[12]. ANOSOV, D. V., Structural stability and ergodicity of geodesic flows on compact, Riemannian manifolds of negative curvature. (To appear.)

[13]. ARNOLD, V. I. & SINAI, Y. G., *Dokl. Akad. Nauk S.S.S.R.*, 144, 4 (1962), 695.

[14]. SMALE, S., *Report on the Symposium on Non-linear Oscillations.* Kiev Math. Institute, 1961.

SOME QUESTIONS OF UNIFORM TOPOLOGY

By JU. SMIRNOV

My report contains three parts: the proximity, the uniformity and the general metrics. I should like to tell you about some themes which are interesting for Moscow's topologists. I should like to suggest a few open questions also.

I. Proximity

The first axioms of the proximity were given by F. Riesz at the mathematical congress in Bologna in 1908. He gave the definition of general proximity-space. The complete axiomatic was given by W. Ephremovitsch [1] in 1935. He added a separation axiom. With this axiom the apparatus of real-valued functions and pseudometrics can be applied. I proved that the notion of proximity-space is closely connected and is indeed equivalent to the notion of compactification. Thus the proximity-theory can be considered as a part of the theory of topological spaces. In fact, every compact **C** has one and only one proximity:

$$A\delta B(^1) \quad \text{if and only if} \quad \bar{A}^c \cap \bar{B}^c \neq \phi,(^2) \tag{1}$$

$$\text{where} \quad A \subseteq C, B \subseteq C.$$

Therefore every compactification **C** of the given space **X** defines this natural proximity (1) on the X. The received correspondence between all compactifications of a given space X and all its proximities is one-to-one and onto [2]. Thus every proximity-space P can be defined as a completely regular space considered together with some compactification u**P**. From this point of view the bounded real-valued function g on the **P** is proximity-continuous if and only if it has a continuous extension on the u**P**. Moreover the map f of the proximity-space X into the proximity-space Y is proximity coninuous if and only if it has a continuous extension f mapping the compactification u**X** into u**Y** [2]. These facts have been reproved by many authors for example [3], [10].

Therefore we have connection between the proximity-spaces and subrings of the ring $C(X)$ of all bounded real-valued continuous functions. The classical Weierstrass-Stone's theorem for proximity-spaces is a consequence of this situation [4].

The metric spaces have the following natural proximity:

$$A\delta B \quad \text{if and only if} \quad \varrho(A,B)=0.(^3) \tag{2}$$

I note that the notion of proximity-continuity and the notion of uniform continuity are equivalent for the mappings of the metric spaces. For

(1) $A\ \delta\ B = A$ is near to B.
(2) $\bar{A}^c$ is the closure of A.
(3) $\varrho(A,B) = \inf \varrho(x,y)$, where $x \in A$, $y \in B$ and $\varrho(x,y)$ is the given metric.

proximity-spaces I have given notions of completeness, of the completion and of full-boundness [5]. These notions are founded on the consideration of some kind of coverings or of pseudometrics [6]. All these notions are equivalent to the corresponding metrical notions for metric spaces considered with the natural proximity (2). We can get the completion $c\mathbf{P}$ of the given metric space P from the corresponding compactification $u\mathbf{P}$ by removing all the points with first countable-axiom [2]. Therefore, the metric space P is complete if and only if no one point of the difference $uP \setminus P$ satisfies the first countable-axiom. Thus some uniform properties of the metric space P can be characterised by topological properties of the difference $uP \setminus P$. The second example: The metric space $\mathbf{P}$ is full-bounded in usual sense if and only if the difference $uP \setminus P$ is hereditary-normal [17]. I note that not all uniform properties of the metric space $\mathbf{P}$ can be characterized in this manner (for example, the property being zero-uniform-dimensional).

In topological groups there are also a natural proximity:

$A\delta B$ if and only if $A \cap O \cdot B \neq \phi$ for every neigbourhood O of the unity.(1) (3)

The proximity of groups was studied very little.

QUESTION 1 (A. N. Kolmogoroff). Is it possible to define the completeness of a topological group P using only the properties of this natural proximity? In other words: are there proximity-homeomorphic groups G and H where G is complete and H is not?

QUESTION 2. Are there uniform properties of topological groups which can not be characterized by proximity-properties?

I note that for metric spaces each uniform property is a proximity-property. It follows that all the theorems of the uniform topology of metric spaces can be formulated and proved in terms of proximity.

II. Uniformity

The theory of uniform spaces has been constructed by A. Weil in 1935 earlier than the theory of proximity [7]. The uniform space can be defined by many ways: the Bourbaki's way [8] by a certain system of neighbourhoods of diagonal of square $X \times X$, the Tuckey's way [9] by a certain system of coverings of X, the very general Császàr's way [10] by a certain ordering of set of all subsets of X. The natural notions of completeness and full-boundness (precompactness) are equivalent to the corresponding metric properties for metric space.

For uniform spaces there are given generalizations of some theorems of analysis (for instance, theorem of Ascoli). The apparatus of uniform topology is very convenient for linear topologic spaces [11].

Every uniform space U has a natural proximity: Let the uniformity of topological space X be defined by the structure $\Sigma = \{\omega_\lambda\}$ of coverings ω_λ. Then

$A\delta B$ if and only if each covering ω_λ has an element O_λ that $A \cap O_\lambda \neq \phi$ and $B \cap O_\lambda \neq \phi$. (4)

(1) We can get the other proximity putting $A \cap B.O \neq \phi$ for non-commutative groups.

Let P be some proximity-space and P_α is some uniform space on the set P for which the proximity (4) coincide with the proximity of P. Let Σ_α be the uniform structure of the coverings which defines the uniform space P_α. In this situation the completion cP_α of the space P_α is the maximal among those subsets of the compactification uP on which every open covering of the structure Σ_α can be extended into an open covering [5, 12]. Thus we have a well defined map of the set of all uniform spaces P_α with the same proximity P into the set of all subsets of the compactification uP.

But this map is no one-to-one even for complete uniform spaces. On the set of all natural numbers one can define many (continuum!) uniform spaces where every such space is complete and proximity-discrete. Moreover on the set M of the cardinality of the continuum one can construct hypercontinuum complete uniform spaces with discrete proximity where every such space has the same set of all uniform-continuous real-valued functions. Thus the uniform space U can not be in general characterised by the group $C'(U)$ of all uniform continuous real-valued functions.

There is the minimal space among all uniform spaces U_α defined on the same proximity-space P and having the same group $C'(U_\alpha)$ (i.e. all these groups are naturally isomorphic). These minimal spaces are defined in another way by J. Isbell [13]. He calls these spaces weak-uniform. The natural map of the set of all weak-uniform spaces (defined on the same topological space X) into the set of all subgroups of the group $C'(X)$ of all real-valued continuous functions is one-to-one.

Let $C'(X)$ be the ring of all real-valued continuous functions on the topological space X and $C'(P)$ the group of all real-valued proximity-continuous functions on the proximity-space P.

QUESTION 3. Is it possible to characterize groups $C'(U_\alpha)$ where U_α are the uniform spaces on X (or on P) among all subgroups of the group $C'(X)$ (or corresponding of $C'(P)$) by their structural, algebraic and topological properties?

J. Isbell proved the Weierstrass-Stone's theorem for all these weak-uniform spaces U for which the groups $C'(U)$ are algebras having some composition-property. Let us have $P \subseteq S \subseteq uP$ where P is a some proximity-space, uP is its natural compactification and S—some Q-space in the sense of E. Hewitt. Then the group $C'(S)$ is an algebra and it has the composition-property.

QUESTION 4. Are only these groups $C'(S)$ algebras with the composition-property or not?

III. General metrics

At the last year there appeared a new theory of general metric spaces. It is clear that in every such case the distances between other points of a given metric space can be no real numbers—the elements of the real line E'—but elements of some its generalization. This generalization is a notion of a topological semifield given by T. Sarymsacov, V. Boltianski and M. Antonovski [14].

The topological semifield is defined as a commutative and associative ring in which some subset of its elements is chosen. They call these elements

positive and this subset—cone. This cone should have a number of natural properties. It is easy to define the notion of completeness for topological semifields. The fundamental example of complete topological semifield is the product E^α of a finite or infinite number of real lines E^1, where topology is Tychonov topology and the operations are the following:

$$\{x_\alpha\}+\{y_\alpha\}=\{x_\alpha+y_\alpha\},\ \{x_\alpha\}\cdot\{y_\alpha\}=\{x_\alpha\cdot y_\alpha\}. \tag{5}$$

There are a natural ordering: $\{x_\alpha\}\leqslant\{y_\alpha\}$ if $x_\alpha\leqslant y_\alpha$ for every α. The cone of the E^α is a set of all elements $\{x_\alpha\}$ which $>\{o\}$. Thus every euclidean space E^n is a complete topological semifield. It was proved that every complete topological semifield is topological isomorph to some space E^α. By definition and by this proposition the distances $\varrho(x,y\}$ between other points of a given general metric space are the elements of some semifield E^α. Analogically we can define the general normed space: the norm $\|x\|$ of every point x from this space is a element of some semifield E^α. Of course the axioms are the usual axioms of real-valued metric or real-valued norm. The every semifield E^α is a general normed space because we can put that $\|\{x_\alpha\}\|=\{x_\alpha\}$. Thus it is a general metric space with distance $\varrho(\{x_\alpha\},\{y_\alpha\})=\{|x_\alpha-y_\alpha|\}$.

Every general metric space has a natural "metric" uniformity. Let ε be a some neighbourhood of the unity $\{1\}$ of E^α. Then we have the "ε-neighbourhood" $U_\varepsilon x$ for every point x of the general metric space R. This "ε-neighbourhood" is a set of all this points y of R for which $\varrho(x,y)\in\varepsilon$. The set Σ of all coverings of R having a refinement $\omega_\varepsilon=\{U_\varepsilon x, x\in R\}$ for some ε is a uniform structure. The (by other ways defined) notions of the compactness, of the full-boundness, of the completeness for general metric space are equivalent to corresponding notions for this metrical uniformity. It is easy to prove that every uniform space is metrizable by some general metric.

The three authors have proved the Kolmogoroff's theorem on a possibility of introducing the norm for linear topological spaces. They have generalized the Banach's theorem on open mappings on the general normed (and metric) spaces and the theory of abstract ergodic theorems of Eberlein [15].

Thus we see that this new theory of general metric (and normed) spaces is very convenient for analysis (more than uniform spaces and semiordered spaces of Kantorovitsch [18]).

QUESTION 5. Is it possible to make the axiomatic of topological semifields more and essential simple than the given?

QUESTION 6. Is it possible to prove the Banach's theorem on open mappings (see 12.2. from [16]) on the uniform spaces?

REFERENCES

[1]. EPHREMOVITSCH, W., *Dokl. Akad. Nauk SSSR*, 76, N 3 (1951).
—— Mat. Sb., *31*, N 1 (1952).
[2]. SMIRNOV, JU., *Mat. Sb. N.S.*, 31, N 3 (1952).
[3]. LEADER, S., *Fund. Math.*, XLVII (1959).
[4]. SMIRNOV, JU., *Čech. Math. J.*, *10* (1960).
[5]. SMIRNOV, JU., *Trudy Mosk. Matem. Obsh.*, 3 (1954); 4 (1955).
[6]. LEADER, S., *Fund. Math.*, XLIII (1960).

[7]. WEIL, A., *Sur les espaces a structures uniforme*. Paris, 1937.
[8]. BOURBAKI, N., Topologie générale. *Actualités Sci. Ind.* No. 858. Paris, 1951.
[9]. TUCKEY, Y., *Convergence and Uniformity in Topology*. Princeton, 1940.
[10]. CSÁSZÁR, A. *Fondements de la topologie générale*. Budapest, 1960.
[11]. NAKANO, H, *Linear Topological Spaces*. Tokyo, 1951.
[12]. SHIROTA, T., *Osaka Math. J.*, 2, No. 2 (1950).
[13]. ISBELL, J., *Ann. Math.*, 68, No. 1 (1958).
[14]. ANTONOVSKI, M., BOLTIANSKI, V. & SARYMSAKOV, T., *Topologitscheskie polupolia*. Tashkent, 1960.
[15]. SARYMSAKOV, T., Report on the *Internat. Congress Math.*, 1962.
[16]. ANTONOVSKI, M., BOLTIANSKI, V. & SARYMSAKOV, T., *Meritscheskie prostranstva nad polupoliami*. Tashkent, 1961.
[17]. OREVKOV, JU., *Dokl.*, *Akad. Nauk SSSR*, 143, N 2 (1962).
[18]. KANTOROVITSCH, L., VULICH, B. & PINSKER, A., *Funktionalny analys v poluuporiadotschennych prostranstvach*. Moscow–Leningrad, 1950.

SOME ASPECTS OF STABLE HOMOTOPY THEORY

By GEORGE W. WHITEHEAD

1. The suspension category

Many of the phenomena of homotopy theory become simpler in the "suspension range". This fact led Spanier and J. H. C. Whitehead in 1953 [26] to introduce the suspension category $\mathcal{S}$ and to propose the study of its homotopy theory as a first approximation to the homotopy theory of the category of spaces. The objects of $\mathcal{S}$ are spaces, but its maps are equivalence classes of maps: $S^p X \to S^q Y$, two such maps being identified if their iterated suspensions of some order are homotopic. Thus the set $\{X, Y\}$ of S-maps of X into Y acquires the structure of a graded abelian group, whereas the set $[X, Y]$ of homotopy classes of maps of X into Y has no particular structure.

In subsequent years, Spanier and Whitehead pursued this program in a series of papers [27, 28, 29], culminating in the *Spanier-Whitehead duality*. This duality asserts the existence of a natural isomorphism $\{X, Y\} \approx \{D_n Y, D_n X\}$, where X, Y are subcomplexes of S^n and $D_n X, D_n Y$ are essentially their complements (a more general formulation has been given by Spanier in [25]). Their duality has been applied to a variety of problems, e.g. to the problem of imbeddibility of a complex in S^n [24].

2. Spectra and homology theory

For many purposes the suspension category has proved to be too "small"; a more suitable category seems to be the category $\mathcal{S}p$ of *spectra*. These were introduced by Lima in [20] and have since been studied, under a variety of restrictions, in [1, 25, 31]. A *spectrum* $\mathbf{E}$ in the most general sense is simply a sequence of spaces E_n, together with a sequence of maps $f_n: SE_n \to E_{n+1}$ (or, equivalently, a sequence of maps $g_n: E_n \to \Omega E_{n+1}$); $\mathbf{E}$ is said to be an Ω-spectrum if and only if each map g_n is a homotopy equivalence. Examples are the *sphere spectrum* $\mathbf{S} = \{S^n\}$ and the *Eilenberg-MacLane spectrum* $\mathbf{K}(\Pi) = \{K(\Pi, n)\}$.

The category $\mathcal{S}p$ seems to provide the proper setting for cohomology operations of higher kind [23]. Moreover, it is closely related to homology and cohomology theories of a very general sort.

The axiomatic study of homology and cohomology theory was inaugurated by Eilenberg and Steenrod in 1945 [11, 12]. Of their axioms, the first six are of a very general character; on the other hand, their seventh, the *dimension axiom*, plays a normative role, singling out the classical theories among the multitude of possible theories which satisfy the first six. Examples of such theories are the stable homotopy and cohomotopy groups, the bordism and cobordism groups studied by Atiyah [2] and by Conner and Floyd [7], and the groups of stable vector bundles studied by Atiyah and Hirzebruch [3, 4], Bott [5] and others.

General cohomology theories have become reasonably familiar in recent years (cf. the work of Eckmann and Hilton [10]). For any space Y, the functor $[\ ,Y]$ of homotopy classes of maps into Y is, in some sense, an "unstable" cohomology theory, and Brown in [6] has given an axiomatic characterization of such theories. One obtains a cohomology theory satisfying the first six axioms of Eilenberg and Steenrod by replacing Y by a spectrum $\mathbf{E}$ and defining, for any compact polyhedron X,

$$H^n(X;\mathbf{E}) = \{X, \mathbf{E}\}^n = \lim_k [S^k X, E_{n+k}].$$

This can be given an alternative description as follows: the function spaces $F(X, E_n)$ of all maps of X into E_n are the components of a spectrum $\mathbf{F}(X,\mathbf{E})$, and

$$H^n(X;\mathbf{E}) \approx \pi_{-n}(\mathbf{F}(X;\mathbf{E})).$$

In [6], Brown has shown that, under mild restrictions, every theory satisfying the first six axioms can be obtained in this way, from an essentially unique Ω-spectrum $\mathbf{E}$. For example, $H^n(X;\Pi) \approx H^n(X;\mathbf{K}(\Pi))$ for any abelian group Π.

On the other hand, homology theories in a general sense have until recently seemed somewhat mysterious. In [31] I proposed the following definition. If X is a compact polyhedron and $\mathbf{E}$ a spectrum, then the reduced joins $X \wedge E_n$ are the components of a spectrum $X \wedge \mathbf{E}$; let

$$H_n(X;\mathbf{E}) = \pi_n(X \wedge \mathbf{E}).$$

One then obtains a homology theory satisfying the first six axioms; moreover

$$H_n(X;\Pi) \approx H_n(X;\mathbf{K}(\Pi)),$$

while $H_n(X;\mathbf{S})$ is the nth stable homotopy group $\Sigma_n(X)$.

3. Duality theorems

Classically, homology and cohomology groups are related by the Alexander and Poincaré duality theorems, and by the universal coefficient theorem. Universal coefficient theorems for general theories have been studied by D. S. Kahn in his thesis (unpublished); the situation is much more complex than in the classical case, a short exact sequence being replaced by a spectral sequence.

The isomorphisms of the classical duality theorems are described by certain products, which are induced by a pairing of two coefficient groups to a third. Using an appropriate notion of pairing of spectra, one can define cup, cap, etc. products which have properties completely analogous to the classical ones. Using these, I have shown in [31] that the analogue of the Alexander duality theorem holds in full generality: for any spectrum $\mathbf{E}$, we have the isomorphism

$$\tilde{H}^q(X;\mathbf{E}) \approx \tilde{H}_{n-q-1}(D_n X;\mathbf{E}).$$

It then follows that the analogue of Brown's theorem holds for homology.

The Poincaré duality theorem is much more delicate. Consider a spectrum

$\mathbf{R}$ which is ring-like in the sense that there is given a self-pairing $(\mathbf{R},\mathbf{R})\to\mathbf{R}$ and a map $\mathbf{S}\to\mathbf{R}$ with suitable properties. An $\mathbf{R}$-*module* $\mathbf{M}$ is then a spectrum with a suitable pairing $(\mathbf{R},\mathbf{M})\to\mathbf{M}$. If $\mathbf{M}$ is an $\mathbf{R}$-module, the cap-product is a pairing

$$H_n(X;\mathbf{R})\otimes H^q(X;\mathbf{M})\to H_{n-q}(X;\mathbf{M}).$$

If X is a compact triangulable manifold and $z\in H_n(X;\mathbf{M})$ a suitably defined fundamental class then

$$z\cap : H^q(X;\mathbf{M})\to H_{n-q}(X;\mathbf{M})$$

is an isomorphism. However, the question of existence of a fundamental class is extremely complex. For example, if $\mathbf{R}=\mathbf{S}$, then X has a fundamental class if and only if X is a *homotopy* Π-*manifold*; for the differentiable case this means that the stable normal bundle is fibre-homotopically trivial [21]. Thus the homotopy Π-manifolds are precisely those manifolds for which Poincaré duality holds for arbitrary spectra.

4. Semi-simplicial spectra

Semi-simplicial complexes were introduced by Eilenberg and Zilber in [31]. In recent years, they have been systematically exploited by Kan [15, 16, 17], Moore [22], Dold [8] and others, and this work has been responsible for much of the great advance that has been made by algebraic topology in recent years. One advantage of this approach is that one can apply to a semi-simplicial complex a functor from the category of sets to a category with a richer algebraic structure, and then to use algebraic machinery to study the resulting object. An example is the F-construction of Milnor (unpublished); $F(K)$, the free group complex generated by K, is the semi-simplicial analogue of the space of loops on the suspension of K; thus the map $K\to FK$ is a homotopy equivalence in the stable range.

Recently Kan [18] has introduced semi-simplicial spectra. For these the functor F is a weak homotopy equivalence; thus, to all intents and purposes and spectrum $\mathbf{E}$ can be replaced by $F(\mathbf{E})$.

In recent unpublished work, Kan and I have applied the semi-simplicial techniques to the question of Poincaré duality. On the one hand, there is a classification of spectra by "degrees of abelianness"; an abelian spectrum is essentially a product of Eilenberg-MacLane spectra, and a spectrum is n-abelian if it is dominated by an n-stage Postnikov extension whose factors are abelian. On the other hand, there is a classification of manifolds by "degrees of orientability"; a manifold is 1-orientable if and only if it is orientable in the usual sense, and n-orientable if and only if all stable homology operations of the $(n-1)$st kind vanish on the fundamental class. It turns out that a manifold is n-orientable if and only if it satisfies Poincaré duality for all n-abelian spectra. Moreover, there exist n-orientable manifolds for some $n, 2\leqslant n<\infty$.

These results were obtained by studying the functor which associates to every group its commutator subgroup. This functor is one of a class of "fully invariant" functors, determined by the fully invariant subgroups of the free group on countably many generators [19, § 37]. If T is such a functor, then $\mathbf{R}(T)=F(\mathbf{S})/TF(\mathbf{S})$ is a ring-like spectrum in the sense

explained above, and the $\mathbf{R}(T)$-mdoules are those spectra which are dominated by $F(\mathbf{E})/TF(\mathbf{E})$ for some spectrum $\mathbf{E}$. One then has the classes of $\mathbf{R}(T)$ orientable manifolds; the relationships among these classes remain an object for future study.

5. Conclusion

This lecture has not aimed at completeness. Some of the most interesting developments have been reported on by others at this Congress. It may be remarked that most of this work has examined stable homotopy theory for its own sake. On the other hand, the viewpoint of Spanier and Whitehead toward stable homotopy as a first approximation to homotopy theory suggests the possibility of higher-order approximations. There are a few examples of "semi-stable" phenomena. Among them are the exact sequence involving the homology suspension which I found in [30]; the exact sequence of James [14] involving the homotopy suspension; and the analogous exact sequences studied by Dold and Puppe in [9]. In this connection, the special role played by quadratic functors in [9] is suggestive.

REFERENCES

[1]. Adams, J. F., Theorie de l'homotopie stable. *Bull. Soc. Math. France*, 87 (1959), 277–280.

[2]. Atiyah, M. F., Bordism and cobordism. *Proc. Cambridge Philos Soc.*, 57 (1961), 200–208.

[3]. Atiyah, M. F. & Hirzebruch, F., Riemann-Roch theorems for differentiable manifolds. *Bull. Amer. Math. Soc.*, 65 (1959), 276–281

[4]. —— Vector bundles and homogeneous spaces. *Proc. Symp. Pure Math.*, Vol. III, 7–37. Providence, 1961.

[5]. Bott, R., Quelques remarques sur les théorèmes de périodicité. *Bull. Soc. Math. France*, 87 (1959), 293–310.

[6]. Brown, E. H., Cohomology theories. *Ann. Math.*, 75 (1962), 467–484.

[7]. Conner, P. E. & Floyd, E. E., Differentiable periodic maps, *Bull. Amer. Math. Soc.*, 68 (1962), 76–86.

[8]. Dold, A., Homology of symmetric products and other functors of complexes. *Ann. Math.*, 68 (1958), 54–80.

[9]. Dold, A. & Puppe, D., Homologie nicht-Addiviter Funktoren. *Ann. Inst. Fourier*, 11 (1961), 201–312.

[10]. Eckmann, B. & Hilton, P. J., Groupes d'homotopie et dualité, *C. R. Acad. Sci. Paris*, 246 (1958), 2444–2447, 2555–2558, 2991–2993.

[11]. Eilenberg, S. & Steenrod, N. E., Axiomatic approach to homology theory. *Proc. Nat. Acad. Sci. U.S.A.*, 31 (1945), 117–120.

[12]. —— Foundations of Algebraic Topology. Princeton Univ. Press, 1952.

[13]. Eilenberg, S. & Zilber, J. A., Semi-simplicial complexes and singular homology. *Ann. Math.*, 51 (1950), 499–513.

[14]. James, I. M., On the suspension triad. *Ann. Math.*, 63 (1956), 191–247.

[15]. Kan, D. M., On c.s.s. complexes. *Amer. J. Math.*, 79 (1957), 449–476.

[16]. —— A combinatorial definition of homotopy groups. *Ann. Math.*, 67 (1958), 282–312.

[17]. —— On homotopy theory and C.S.S. groups. *Ann. Math.* 68 (1958), 38–53.

[18]. —— Semi-simplicial spectra. *Illinois J. Math.* (To appear.)

[19]. KUROSH, A. G., *The Theory of Groups*. New York, 1956.

[20]. LIMA, E. L., The Spanier-Whitehead duality in new homotopy categories. *Summa Bras. Math.*, 4 (1958–9), 91–148.

[21]. MILNOR, J. & SPANIER, E. H., Two remarks on fibre homotopy type. *Pacific J. Math.*, 10 (1960), 585–590.

[22]. MOORE, J. C., Semi-simplicial complexes and Postnikov systems. *Symp. Internacional Topologia Algebraica*. Mexico City, 1958, 232–247.

[23]. PETERSON, F. P., Functional cohomology operations. *Trans. Amer. Math. Soc.*, 86 (1957), 197–211.

[24]. —— Some non-embedding problems. *Bol. Soc. Mat. Mexicana*, 1957, 9–15.

[25]. SPANIER, E. H., Function spaces and duality. *Ann. Math.*, 70 (1959), 338–378.

[26]. SPANIER, E. H. & WHITEHEAD, J. H. C., A first approximation to homotopy theory. *Proc. Nat. Acad. Sci. U.S.A.*, 39 (1953), 655–660.

[27]. —— Duality in homotopy theory. *Mathematika*, 2 (1955), 56–80.

[28]. —— The theory of carriers and S-theory. *Algebraic Geometry and Topology*, 330–360. Princeton, 1957.

[29]. —— Duality in relative homotopy theory. *Ann. Math.*, 67 (1958), 203–238.

[30]. WHITEHEAD, G. W., On the homology suspension. *Ann. Math.*, 62 (1955), 254–268.

[31]. —— Generalized homology theories. *Trans. Amer. Math. Soc.*, 102 (1962), 227–283.

ON RESOLUTION OF SINGULARITIES (CHARACTERISTIC ZERO)(1)

By HEISUKE HIRONAKA

§ 1. Main results and applications

Given a complex-analytic *variety* V (i.e. a reduced complex-analytic space in the sense of H. Grauert), the problem of *resolution of singularities of* V is to find a proper modification $f: \tilde{V} \to V$ such that the modified variety $\tilde{V}$ is free from singularities, i.e. $\tilde{V}$ is a complex-analytic *manifold*. Here we use the term "modification" to signify a morphism (holomorphic map), say $f: \tilde{V} \to V$, which induces an isomorphism (biholomorphic homeomorphism) of a dense open subset of $\tilde{V}$ onto such an open subset of V. We see that if a modification $f: \tilde{V} \to V$ is *proper* (i.e. the preimage by f of every compact subset of V is compact) then f induces an isomorphism of $\tilde{V} - \tilde{W}$ onto $V - f(\tilde{W})$ where $\tilde{W}$ is an analytic set in $\tilde{V}$ which is nowhere dense. The above problem is still open except in lower dimensions,(2) but the results

(1) On resolution of singularities of an algebraic surface, various proofs have been proposed by Italian geometers, such as P. del Pesso, B. Levi, Severi, Albanese, Chisini. An excellent description of their basic ideas towards resolution can be found in Zariski's book [4] and, in spite of criticism by minds of rigor, one can perceive the right philosophy which leads to a satisfactory solution of the problem. As a matter of fact, the most basic idea that underlies our inductive proof of resolution in all dimensions has its origin in B. Levi's works [1,2], or more precisely, in the theorem of Beppo Levi which was clearly stated and given a rigorous proof by O. Zariski in his paper [5]. This theorem asserts, at least, our Theorem I of the first section for the case in which V is a hypersurface on a non-singular algebraic 3-dimensional manifold. A rigorous proof of resolution of singularities of an algebraic surface was first given by J. Walker, *Ann. Math.*, 36 (1935) and then by O. Zariski [3] by purely algebraic methods. For an algebraic variety of dimension 3, it was given by O. Zariski [5]. This proof of Zariski in dimension 3 heavily depends upon the local uniformization theorem for a function field of transcendence degree 3. Although the local uniformization theorem is proved in all dimensions, by Zariski himself [6], it seems more difficult to find an essential use of it in higher dimensions than to prove the resolution directly in all dimensions as is done in this paper, although the passage from local uniformization to global resolution was proved surprisingly short for surfaces and successfully made up in dimension 3 with help of a stronger resolution theorem for embedded surfaces. (See Zariski, *Ann. Math.*, 62 (1942), and also [5].)

(2) Some of the classical methods of resolving singularities of an *algebraic* surface can be extended to the resolution of singularities of a *complex-analytic* surface in the sense that stated in the beginning of this section. For instance, rather simple modification of Zariski's method in [3] with reference to an ideal theory in a regular local ring of dimension two proves it possible to resolve singularities of a complex-analytic surface by alternatively repeated normalizations and quadratic transformations. As for the case of dimension 3, Kuhlmann (Universität Würzburg) has succeeded in proving resolution of singularities of a complex-analytic variety which admits a finite morphism onto a complex-anlytic 3-dimensional manifold. We note here, however, that we can generalize the theorem of Beppo Levi to eliminate singularities of codimensions $\leqslant 2$ on an arbitrary complex-analytic variety and, combining it with our results in the first section, prove the resolution of singularities for a complex-analytic variety of dimension $\leqslant 3$ under the single condition that the topology of variety is countable at infinity.

of this work include a satisfactory answer to the problem for certain restricted classes of complex-analytic varieties, such as algebraic varieties, Stein varieties, etc., in all dimensions. To be precise, we define and denote by $\mathcal{A}_V$ what we call *GAGA structure on a complex-analytic variety* V as follows: Let $\mathcal{O}_V$ denote the structural sheaf of local rings of V (i.e. the sheaf of germs of holomorphic functions on V). Then a *GAGA structure on* V is, by definition, a coherent sheaf of subrings of $\mathcal{O}_V$ (coherent as a sheaf of rings in the sense of J. P. Serre) which satisfies the following conditions:

(i) *For every point x of V, there exists a coordinate system of V in a neighborhood of x such that the germs of coordinate functions belong to the stalk $\mathcal{A}_{V,x}$ of $\mathcal{A}_V$ at x.*

(ii) *$\mathcal{O}_V$ is faithfully flat over $\mathcal{A}_V$*, i.e. *$\mathcal{O}_{V,x}$ (the stalk of $\mathcal{O}_V$ at x) is so over $\mathcal{A}_{V,x}$ for every x.*

(iii) *For every x, every ideal in $\mathcal{A}_{V,x}$ is the stalk of a coherent sheaf of ideals in $\mathcal{A}_V$ (global on V).*

(iv) *If S is an $\mathcal{A}_{V,x}$-algebra of finite type, then the singular locus of the scheme* Spec $(S\otimes_{\mathcal{A}_{V,x}}\mathcal{O}_{V,x})$ *is the preimage of that of* Spec (S) *by the canonical morphism of the first scheme to the second.*

We remark that, in virtue of (ii), the subring $\mathcal{A}_{V,x}$ of $\mathcal{O}_{V,x}$ is *local* and *noetherian* for all x and that a sheaf of $\mathcal{A}_V$-modules N is coherent if and only if the sheaf of $\mathcal{O}_V$-modules $N\otimes_{\mathcal{A}_V}\mathcal{O}_V$ is so. A point of a scheme is said to be *simple* (or *multiple*) if the local ring at the point is regular (resp. not regular and the *singular locus* of a scheme is, by definition, the set of multiple points of the scheme. In view of (i) and (ii), (iv) is equivalent to saying that the singular locus of Spec $(S\otimes_{\mathcal{A}_{V,x}}\hat{\mathcal{A}}_{V,x})$ is the preimage of that of Spec(S) by the canonical morphism, where $\hat{\mathcal{A}}_{V,x}$ denotes the completion of the local ring $\mathcal{A}_{V,x}$. It follows that the singular locus of Spec(S) is closed in the topology of scheme (i.e. Zariski topology) for every $\mathcal{A}_{V,x}$ algebra S of finite type. The above four conditions on a sheaf of subrings $\mathcal{A}_V$ of $\mathcal{O}_V$ will be referred to as *GAGA conditions.*

The primary purpose of this work is to prove the resolution of singularities of an arbitrary complex-analytic variety which admits a GAGA structure and whose topology is *countable at infinity* in the sense that the underlying topological space is a union of countable compact subsets. First we give some examples of complex-analytic varieties satisfying these conditions.

Example I. *Algebraic varieties defined over fields of characteristic zero.* We say that a complex-analytic variety V is *algebraic* (or *an algebraic variety*), if there is given a finite open covering $\{U_i\}_{i\in I}$ of V together with an isomorphism h_i of each U_i (as a complex-analytic variety) to a complex-analytic subvariety of a complex number space defined by a finite system of polynomial equations such that in each of the intersections $U_i\cap U_j$ the coordinate transformations with reference to h_i and h_j are rational. We say that an algebraic variety V is *defined over a field k* (a subfield of the field of complex numbers), if the field k contains all the coefficients appearing in the polynomial equations for U_i and in the rational transformations in $U_i\cap U_j$. If this is the case, we can speak of a *rational function on V defined over the field k*, that is, a meromorphic function on V whose restriction to each U_i is induced by a rational function with coefficients in k on the ambient complex number space of $h_i(U_i)$. Let $\mathcal{A}_V$ be the sheaf of germs of holomorphic rational functions defined over k on V, which is a coherent sheaf of subrings

of O_V and satisfies the GAGA conditions defined above. We shall call $\mathcal{A}_V$ an *algebraic structure on* V (*defined over the field* k).

Example II. *Stein varieties*. A complex-analytic variety V is called a *Stein variety*, if its topology is countable at infinity and "Theorems A and B" hold for every coherent sheaf of O_V-modules N, i.e. for each $x \in V$ the image of $H^0(V,N)$ in N_x (the stalk of N at x) generates N_x as $O_{V,x}$-module and $H^q(V,N)$ vanishes for all positive q. Now let $A = H^0(V,O_V)$, that is, the ring of holomorphic functions on V. If x is any point of V, we can associate with it a maximal ideal $m(x)$ in A which consists of those functions in A which vanish at x. The ring of fractions $A_{m(x)}$ is then a subring of $O_{V,x}$, and we obtain a unique subsheaf $\mathcal{A}_V$ of O_V such that $\mathcal{A}_{V,x} = A_{m(x)}$ for all x. One can prove that this sheaf of subrings of O_V is coherent and satisfies the GAGA conditions defined above.

Example III. Let V be a complex-analytic variety with a GAGA structure $\mathcal{A}_V \subset O_V$. Let $\mathfrak{J}$ be a coherent sheaf of ideals in $\mathcal{A}_V$ on V, and $\mathfrak{J}O_V$ the sheaf of ideals in O_V generated by $\mathfrak{J}$ which is necessarily coherent. We see that the sheaf of quotient rings $\mathcal{A}_V/\mathfrak{J}$ has no nilpotent elements if and only if $O_V/\mathfrak{J}O_V$ does not. This being assumed, we have a complex-analytic subvariety W of V such that $\mathfrak{J}O_V$ is the sheaf of germs of holomorphic functions on V which vanish on W, so that we may identify $O_V/\mathfrak{J}O_V$ with the structural sheaf of local rings on W. Moreover, we see that the canonical homomorphism of $\mathcal{A}_V/\mathfrak{J}$ into $O_V/\mathfrak{J}O_V$ is injective. Finally, $\mathcal{A}_V/\mathfrak{J}$ viewed as a sheaf of subrings of $O_V/\mathfrak{J}O_V$ satisfies the GAGA conditions. A complex-analytic subvariety W of V will be called a *GAGA subvariety of* V, if it is obtained as above, and the GAGA structure $\mathcal{A}_V/\mathfrak{J}$ on W will be called the *induced GAGA structure on* W.

Example IV. Let V be a complex-analytic variety, and $\mathfrak{J}$ a coherent sheaf of non-zero ideals in O_V. Then there exists a modification $f: \tilde{V} \to V$ (unique up to an isomorphism) associated with $\mathfrak{J}$ by the following universal mapping property:

(*a*) *The inverse image* $f^{-1}(\mathfrak{J})$ (*i.e. the coherent sheaf of ideals in* $O_{\tilde{V}}$ *generated by* $\mathfrak{J}$) *is invertible* (*i.e. locally isomorphic to* $O_{\tilde{V}}$ *as* $O_{\tilde{V}}$*-modules*).

(*b*) *If* $f': \tilde{V}' \to V$ *is any modification such that* $f'^{-1}(\mathfrak{J})$ *is invertible, then there exists a unique morphism* $g: \tilde{V}' \to \tilde{V}$ *such that* $f \circ g = f'$.

We shall call the above f the *$\mathfrak{J}$-modification of* V. In particular, when $\mathfrak{J}$ is the sheaf of ideals of a complex-analytic subvariety B in V, the $\mathfrak{J}$-modification of V will be called *the monoidal transformation of* V *with center* W. We shall be particularly interested in the case when $\mathfrak{J}$ is generated by a sheaf of ideals $\mathfrak{J}$ in $\mathcal{A}_V$, if a GAGA structure $\mathcal{A}_V$ on V is given. Suppose this be the case, and let $f: \tilde{V} \to V$ be the $\mathfrak{J}(=\mathfrak{J}O_V)$-modification of V. Then one can prove that *there exists a unique GAGA structure* $\mathcal{A}_{\tilde{V}}$ *on* $\tilde{V}$ *such that the morphism* f *induces an epimorphism of the ring of fractions of* $\mathcal{A}_{V,f(\tilde{x})}$ *onto that of* $\mathcal{A}_{\tilde{V},\tilde{x}}$ *for every* $x \in \tilde{V}$. We shall simply call $\mathcal{A}_V$ *the canonical GAGA structure on* $\tilde{V}$ whenever the context makes the sense clear.

Let S be a commutative ring with unity, M an S-module and I an ideal in S. Then we consider the S/I-modules $gr_I^n(M) = I^nM/I^{n+1}M$ for all non-negative integers n, and the direct sum $gr_I(M) = \sum_{n=0}^{\infty} gr_I^n(M)$. In particular $gr_I(S)$ has a canonical structure of graded S/I-algebra, and in general $gr_I(M)$ a canonical structure of graded $gr_I(S)$-module. We globalize this

definition as follows: Let V be a complex-analytic variety (or a scheme in the sense of Grothendieck) whose structural sheaf of local rings is denoted by O_V. Let B be a complex-analytic subvariety (resp. a subscheme) (1) of V, and $\mathfrak{J}$ the sheaf of ideals of B on V. If $\mathfrak{M}$ is a coherent sheaf of O_V-modules on V, the quotient sheaf $\mathfrak{J}^n\mathfrak{M}/\mathfrak{J}^{n+1}\mathfrak{M}$ can be viewed as a sheaf of O_B ($=O_V/\mathfrak{J}$ restricted to B)-modules and is coherent. We denote it by $gr_B^n(\mathfrak{M})$ and set $gr_B(\mathfrak{M})=\sum_{n=0}^{\infty} gr_B^n(\mathfrak{M})$. In particular, we write $gr_B(V)$ for $gr_B(O_V)$, which is a quasi-coherent sheaf of graded O_B-algebras.

DEFINITION 1. *Let V and B be as above. Then we say that V is normally flat along B, if $gr_B(V)$ is flat over O_B, or equivalently, $gr_B^n(V)$ is locally free as a sheaf of O_B-modules for all non-negative integers n.*

Remark. Suppose V be a complex-analytic variety with a GAGA structure $\mathcal{A}_V$, and $\mathfrak{J}=\mathfrak{I}O_V$ with a coherent sheaf of ideals $\mathfrak{I}$ in $\mathcal{A}_V$, where $\mathfrak{J}$ denotes the sheaf of ideals of B on V. Then for each point x of V, $gr_{\mathfrak{J}x}(O_{V,x})$ is flat as $O_{V,x}/\mathfrak{J}_x$-module if and only if $gr_{\mathfrak{I}x}(\mathcal{A}_{V,x})$ is flat as $\mathcal{A}_{V,x}/\mathfrak{I}_x$-module. This follows the assumption that O_V is faithfully flat over $\mathcal{A}_V$. Let V_x denote the scheme $\mathrm{Spec}(\mathcal{A}_{V,x})$, and B_x the subscheme of V_x defined by the ideal $\mathfrak{I}_x$ in $\mathcal{A}_{V,x}$. Then the above fact can be expressed by saying that *V is normally flat along B if and only if V_x is so along B_x for every point x of V.*

DEFINITION 2. *Let V be a complex-analytic variety with a GAGA structure $\mathcal{A}_V$. Then a monoidal transformation $f_1: V_1 \to V$ with center $B \subset V$ will be said to be admissible (for the purpose of resolving singularities of V) if the following conditions are satisfied:*

(i) *B is a GAGA subvariety of V which has no singularities.*

(ii) *B is contained in the singular locus of V.*

(iii) *V is normally flat along B.*

A succession (finite or countably infinite) of monoidal transformations $f_r: V_r \to V_{r-1}$ ($r \geqslant 1$, and $V_0=V$) will be said to be admissible if every one of the f_r is admissible. Here the admissibility is meant with reference to the canonical GAGA structure on V_r obtained inductively starting from the given one on V. (See Example IV.)

We shall say that a succession of monoidal transformations $f_r: V_r \to V_{r-1}$ ($r \geqslant 1$ and $V_0=V$) is *locally finite*, if for every point x of V there exists a neighborhood U of x in V such that, except for a finite number of r, the center of f_r does not meet the preimage of U in V_{r-1}. If this is the case, then we have a proper morphism to V from the inverse limit $\tilde{V}$ of the system $\{f_r\}$ which is also a complex-analytic variety. In fact, we see that the morphism of $\tilde{V}$ onto V is a proper modification. If a GAGA structure is given on V and if all the monoidal transformations in the succession have GAGA subvarieties as their centers, then we get a GAGA structure on $\tilde{V}$ as the direct limit of the canonical GAGA structures on V_r. This GAGA structure $\mathcal{A}_{\tilde{V}}$ on $\tilde{V}$ is uniquely determined by the property that for every point $\tilde{x}$ of $\tilde{V}$ the morphism $f: \tilde{V} \to V$ induces an epimorphism of the ring of fractions of $\mathcal{A}_{V,f(\tilde{x})}$ onto that of $\mathcal{A}_{\tilde{V},\tilde{x}}$. We shall simply call it *the canonical GAGA structure on $\tilde{V}$* whenever the context makes the sense clear. We state the first main result of this work in the following form:

THEOREM I. *Let V be a complex-analytic variety. We assume that a GAGA*

(1) Throughout this paper, we use the term "subscheme" (or "subvariety") to mean "*closed* subscheme" (resp. "*closed* subvariety") unless explicitly otherwise mentioned.

structure is given on V and that the topology of V is countable at infinitey. Then there exists a proper modification $f: \tilde{V} \to V$, obtained by a locally finite admissible succession of monoidal transformations, such that $\tilde{V}$ is a complex-analytic manifold, i.e. it is free from singularities.

Remark 1. Let $s(V)$ denote the singular locus of V. Then the above modification f induces an isomorphism of $\tilde{V} - f^{-1}(s(V))$ onto $V - s(V)$. (See condition (ii) of admissibility.)

Remark 2. If V is given as a complex-analytic subvariety of a complex-analytic manifold X, then the above modification ban be extended in a canonical manner to a proper modification $F: \tilde{X} \to X$, so that $\tilde{X}$ is a complex-analytic manifold containing $\tilde{V}$ as a complex-analytic subvariety, that F induces f in $\tilde{V}$ and that F induces an isomorphism of $\tilde{X} - F^{-1}(s(V))$ onto $X - s(V)$. This comes from the fact that f is obtained by a succession of monoidal transformations with non-singular centers.

Remark 3. Our proof of the theorem is done in such a way that if the GAGA structure on V is an algebraic one then f is obtained by a *finite* admissible succession of monoidal transformations, so that the canonical structure on $\tilde{V}$ is necessarily algebraic (and defined over the same field as the given one). Without referring to the proof, however, we can deduce this fact from the above theorem in view of the fact (due to M. Nagata) that every algebraic variety admits an algebraic compactification, i.e. it can be identified with a Zariski open subset (with the induced algebraic structure) of a complete (i.e. compact as a complex-analytic variety) algebraic variety.

Remark 4. If V is a *projective* algebraic variety, then necessarily so is $\tilde{V}$. In general, if V is a compact complex-analytic variety, then an algebraic structure defined over a given field (a subfield of the field of complex numbers) is unique on V provided it exists. Moreover, if V admits an algebraic structure then every *meromorphic* function on V is a *rational* function defined over the field of complex numbers. The theorem thus implies:

COROLLARY.(1) *If k is any field of characteristic zero and K any field finitely generated over k, then there exists a projective algebraic manifold X, defined over k, such that K is isomorphic to the field of rational functions on X defined over k.*

Let us consider the case in which V is given as a complex-analytic subvariety of a complex-analytic manifold X. As is seen in Remark 2, the modification f of the theorem can be extended in a canonical way to a proper modification $F: \tilde{X} \to X$. The preimage $F^{-1}(V)$ of V in $\tilde{X}$ is then a union of $\tilde{V}$ and the socalled *exceptional hypersurfaces* of $\tilde{X}$ for F, i.e. those irreducible subvarieties of codimension one of $\tilde{X}$ which correspond to subvarieties of codimension greater than one of X. In general, $F^{-1}(V)$ can have complicated singularities though the component $\tilde{V}$ of it does not have any singularities. However, it is possible to choose a different succession of monoidal transformations which gives us a proper modification $F: \tilde{X} \to X$ such that $F^{-1}(V)$ has no other singularities than locally transversal crossings of its irreducible

(1) Without any loss of generality, we may (and do, as a matter of fact) assume that the field k is given as a subfield of the field of complex numbers (and even finitely generated over the field of rational numbers).

components, i.e. *normal crossings* as we define below. We note that if V is connected then necessarily so is $F^{-1}(V)$.

DEFINITION 3. *Let X be a complex-analytic manifold (or a non-singular scheme), E a complex-analytic subvariety (or reduced subscheme) of X everywhere of codimension one, and V a complex-analytic subvariety (or a subscheme) of X. Let x be any point of X. We say that E has only normal crossings at x, if there exists a regular system of parameters $(z_1, z_2, \ldots, z_N)$ of the local ring $O_{X,x}$ of X at x such that every irreducible component of E passing through x is defined by an equation $z_i = 0$ locally at x for some i. We say that E has only normal crossings with V at x, if there exists a regular system of parameters of $O_{X,x}$ which has the same property as above and, at the same time, the property that either V does not pass through x or it is defined by a system of equations $z_i = 0$ locally at x for some i. We shall simply say that E has only normal crossings (or only normal crossings with V) if it is so at every point of X.*

We remark that if E has only normal crossings (with V) then all the irreducible components of E (and V itself) are free from singularities.

THEOREM II. *Let X be a complex-analytic manifold with GAGA structure and V a GAGA subvariety of X. We assume that the topology of V is countable at infinity. Then there exists a proper modification $F: \tilde{X} \to X$ having the following properties:*

(i) *$\tilde{X}$ is a complex-analytic manifold, and F induces an isomorphism of $\tilde{X} - F^{-1}(s(V))$ onto $X - s(V)$, where $s(V)$ denotes the singular locus of V.*

(ii) *If $\tilde{V}$ denotes the unique subvariety of $\tilde{X}$ such that F induces a proper modification $f: \tilde{V} \to V$, and $\tilde{E}$ the closure in $\tilde{X}$ of the complement of $\tilde{V}$ in $F^{-1}(V)$, then $\tilde{E}$ is everywhere of codimension one and has only normal crossings with $\tilde{V}$.*

In fact, the above F can be obtained by a locally finite succession of monoidal transformations $F_r: X_r \to X_{r-1}$ ($r \geqslant 1$, and $X_0 = X$) such that the center B_r of F_{r+1} is a non-singular GAGA subvariety of X_r contained in V_r, that E_r has only normal crossings with B_r and that V_r is normally flat along B_r, for all $r \geqslant 0$, where the preimage of V in X_r is written as the union of V_r and E_r in such a way that the morphism $X_r \to X$ induces a proper modification $V_r \to V$ and maps E_r into $s(V)$. In particular, $\tilde{X}$ has a canonical GAGA structure and $F^{-1}(V)$ (and every one of its components) is a GAGA subvariety of $\tilde{X}$.

COROLLARY 1. *Let X and V be as in the theorem. Then there exists a proper modification $F: \tilde{X} \to X$ such that:*

(i) *X is free from singularities,*

(ii) *F induces an isomorphism of $\tilde{X} - F^{-1}(V)$ onto $X - V$,* and

(iii) *$F^{-1}(V)$ is everywhere of codimension one and has only normal crossings.*

In fact, F can be obtained by a locally finite succession of monoidal transformations with non-singular GAGA subvarieties as centers, so that $\tilde{X}$ has a canonical GAGA structure and $F^{-1}(V)$ is a GAGA subvariety of $\tilde{X}$.

In the above theorem and corollary, if the GAGA structure of X is an algebraic one, then we can choose F to be obtained by a finite succession of monoidal transformations of the type described there, so that the canonical GAGA structure of $\tilde{X}$ is also algebraic and $F^{-1}(V)$ is an algebraic subvariety of $\tilde{X}$.

In view of Nagata's compactification of an algebraic variety, we can deduce:

COROLLARY 3. *Let X be a complex-analytic manifold with algebraic structure. Then there exists an algebraic compactification $\tilde{X}$ of X such that $\tilde{X}$ is a complex-analytic manifold and that $\tilde{X}-X$ is everywhere of codimension one and has only normal crossings.*

Let X be a complex-analytic variety and $\mathfrak{J}$ a coherent sheaf of ideals in O_X. We shall be only interested in the case where no stalk of $\mathfrak{J}$ consists of only zero element. We denote by $c(\mathfrak{J})$ the set of points of X at which $\mathfrak{J}$ is not locally invertible. It can be seen that if $f:X'\to X$ is the $\mathfrak{J}$-modification of X, then $c(\mathfrak{J})$ is equal to the set of points of X at which f^{-1} is not well-defined (as a local holomorphic map). In other words, $c(\mathfrak{J})$ is the smallest closed subset of X such that f induces an isomorphism of $X'-f^{-1}(c(\mathfrak{J}))$ onto $X-c(\mathfrak{J})$. If a GAGA structure $\mathcal{A}_X$ on X is given and $\mathfrak{J}$ is generated by ideals in $\mathcal{A}_X$, hence, by a coherent sheaf of ideals $\mathfrak{I}$ in $\mathcal{A}_X$, then for $\mathfrak{J}_x$ to be invertible it is necessary and sufficient that $\mathfrak{I}_x$ is so, where x is any point of X.

In general, if $f:X'\to X$ is a proper modification of complex-analytic varieties, then the set of those points of X at which f^{-1} is not well-defined is a complex-analytic subvariety of X which is called *the fundamental locus of f in X*. We shall denote this by $c(f)$. We are now interested in the problem of eliminating the fundamental locus by succession of monoidal transformations with non-singular centers, especially for the case in which X is non-singular.

THEOREM III. *Let X be a complex-analytic manifold with GAGA structure $\mathcal{A}_X$, and $\mathfrak{J}$ a coherent sheaf of non-zero ideals on X generated by ideals in $\mathcal{A}_X$. Assume that the topology of $c(\mathfrak{J})$ is countable at infinity. Then there exists a proper modification $f:\tilde{X}\to X$ having the following properties:*

(i) *f is obtained by a locally finite succession of monoidal transformations with non-singular GAGA subvarieties as centers,*

(ii) *f induces an isomorphism of $\tilde{X}-f^{-1}(c(\mathfrak{J}))$ onto $X-c(\mathfrak{J})$, and*

(iii) *$c(f^{-1}(\mathfrak{J}))$ is empty, i.e. $f^{-1}(\mathfrak{J})$ is everywhere invertible, where $f^{-1}(\mathfrak{J})$ denotes the coherent sheaf of ideals on $\tilde{X}$ generated by $\mathfrak{J}$.*

In fact, the succession of monoidal transformations, say $f_r:X_r\to X_{r-1}$ ($r\geqslant 1$, and $X_0=X$), can be so chosen that the center of f_r is contained in $c(\mathfrak{J}_{r-1})$ where $\mathfrak{J}_{r-1}$ denotes the sheaf of ideals on X_{r-1} generated by $\mathfrak{J}$.

If the GAGA structure on X is algebraic, then we can apply the theorem to an algebraic compactification X_1 of X and to an extension of $\mathfrak{J}$ from X to X_1. Hence, in this case, we can choose the succession of monoidal transformations in the theorem to be finite.

Let $f:X'\to X$ be a proper modification of complex-analytic varieties X' and X with algebraic structures $\mathcal{A}_{X'}$ and $\mathcal{A}_X$ respectively. We say that f is *rational* if the homomorphism of O_X into $O_{X'}$, associated with f maps $\mathcal{A}_X$ into $\mathcal{A}_{X'}$, and *birational* if, at the same time, the homomorphism induces an epimorphism of the ring of fractions of $\mathcal{A}_{X,f(x')}$ onto that of $\mathcal{A}_{X',x'}$ for all points x' of X'. From Theorem III, it follows that *if $f:X'\to X$ is birational and X has no singularities, then there exists a finite succession of monoidal transformations with non-singular algebraic centers which gives a proper modification $g:\tilde{X}\to X$ such that $g=f\circ h$ with a birational morphism $h:X\to X'$.* We see that if X' is projective then necessarily so is $\tilde{X}$. In this manner, we can prove:

COROLLARY 1. *Let X be a compact complex-analytic manifold with algebraic structure, i.e. a complete algebraic variety. Then there exists a projective alge-*

braic manifold $\tilde{X}$ with a birational modification $g:\tilde{X}\to X$ which is obtained by a finite succession of monoidal transformations with non-singular centers.

We note that there are many example of compact algebraic manifolds which are non-kahlerian. However, in virtue of the above results, we can see that algebraic non-kahlerian manifolds behaves as well as kahlerian manfiolds with respect to various cohomologies.

For an arbitrary modification of complex-analytic varieties, we can prove the following:

LEMMA. *Let $f:X'\to X$ be a proper modification of complex-analytic varieties, and x any point of X. Then there exists an open neighborhood U of x in X and a coherent sheaf of ideals $\mathfrak{J}$ on U such that if $g:U_1\to U$ denotes the $\mathfrak{J}$-modification of U then we have $g=f\circ h$ with a holomorphic map $h:U_1\to X'$.*

With help of this lemma, we can easily deduce the following fact from Theorem III:

COROLLARY 2. *Let X' and X be complex-analytic manifolds, and $f:X'\to X$ a proper modification. Then the cohomological direct images $R^qf(\mathcal{O}_{X'})$ of the structural sheaf of local rings $\mathcal{O}_{X'}$ of X' are zero sheaves for all $q>0$.*

§ 2. Inductive formulation of the resolution problem(1)

We have stated three main theorems, Theorems I, II and III, in the first section. In this section, we shall reformulate those theorems in such a way that the proof can be carried out by induction. We shall define what we call *RESOLUTION DATA*, i.e. "those to be resolved", which are put into four different types denoted by $(R_I^{N,n}, F)$, $R_I^{N,n}, U)$, (R_{II}^N, F) and R_{II}^N. We shall then define what we mean by "*the resolution data being RESOLVED*". Each resolution datum consists of objects on a non-singular scheme X, called *the ambient scheme* of the datum, and the *RESOLUTION* of the datum will be achieved by a finite succession of certain elementary transformations applied to X. For this purpose we shall define what we call *PERMISSIBLE monoidal transformations of X for a given resolution datum on X*, i.e. the elementary transformations of X which may be used to resolve the datum on X. We shall also define *the TRANSFORM of each datum* on the new ambient scheme obtained by a permissible monoidal transformation for the datum, so that we can give a good sense to speaking of a *SUCCESSION of permissible monoidal transformations* for a given datum. After those preparatory definitions are given, we shall be able to state the four theorems, called Theorems $\mathrm{I}_F^{N,n}$, $\mathrm{I}_U^{N,n}$, II_F^N and II^N, which assert *the resolution of the resolution data* of the types $(R_I^{N,n}, F)$, $(R_I^{N,n}, U)$, (R_{II}^N, F) and R_{II}^N, respectively. The proof of those four theorems can be done altogether by

(1) In this section, we make an essential use of the language of schemes introduced by A. Grothendieck [7]. It is essential firstly for the purpose of generalizing our resolution of singularities to certain complex-analytic varieties, as was described in the first section, and secondly, what is the most important, we can carry out our inductive proof of the resolution theorems stated in this section only when they are formulated for schemes of finite type over certain class of local rings (and not only for those over fields). We must include in our inductive formulation of resolutions, at least "algebraic varieties over formal power series rings (with coefficient fields of characteristic zero)" such as those known as algebroid varieties. This fact can be roughly explained by saying that, according to our proof, certain resolution of *algebraic* singularities requires the same in lower dimensions not only of algebraic singularities but also of *algebroid* singularities derived from them.

certain induction on the pairs of, or single, integers N and n which are attached to the resolution data and accordingly to the resolution theorems. After stating Theorems $\mathrm{I}_F^{N,n}$, $\mathrm{I}_U^{N,n}$, II_F^N and II^N, we shall indicate the diagram of implications among those theorems according to which our inductive proof is carried out.

We begin our preparation towards the statement of the four resolution theorems by restricting our attention once for always to a certain class of schemes, which will be simply called *algebraic schemes* throughout this paper.

In this paper we call a local ring S a *base ring* if it satisfies the following conditions:

(i) *S is a noetherian local ring,*

(ii) *the residue field of S has characteristic zero, and*

(iii) *if A is any S-algebra of finite type and $\hat{S}$ denotes the completion of S, then the singular locus of* $\mathrm{Spec}(\hat{S}\otimes_S A)$ *is the preimage of that of* $\mathrm{Spec}(A)$ *under the canonical morphism of the first spectrum to the second.*

We shall denote by $\mathcal{B}$ the class of *base rings*. Then we can prove that the class $\mathcal{B}$ has the following properties:

(1) *every field of characteristic zero belongs to $\mathcal{B}$,*

(2) *if S belongs to $\mathcal{B}$, then so does the completion of S, and*

(3) *if A is an S-algebra of finite type for $S\in\mathcal{B}$, then all the local rings of* $\mathrm{Spec}(A)$ *belong to $\mathcal{B}$.*

We note that (iii) implies the *closedness* of the singular locus of $\mathrm{Spec}(A)$. Throughout this paper, all the schemes we deal with are *schemes of finite type over base rings* (i.e. over $\mathrm{Spec}(S)$ with $S\in\mathcal{B}$). For the sake of convenience, so far as this paper is concerned, we shall use the term "algebraic scheme" to signify a scheme V which admits a morphism of finite type: $V\to\mathrm{Spec}(S)$ with $S\in\mathcal{B}$. We remark that every local ring of an algebraic scheme belongs to $\mathcal{B}$, and that the singular locus of an algebraic scheme is closed.

Let V be an algebraic scheme, and $\mathcal{O}_V$ the structural sheaf of local rings of V. For a point x of V, $\mathcal{O}_{V,x}$ will denote the stalk of $\mathcal{O}_V$ at x which is often called the local ring of V at x. The *dimension of V at x*, denoted by $\dim_x(V)$, is by definition the Krull dimension of $\mathcal{O}_{V,x}$. If V is not empty, then we define the *dimension of V*, denoted by $\dim(V)$ to be the maximum of $\dim_x(V)$ for all $x\in V$. If W is a non-empty subscheme of V, we define the *codimension of W in V* to be the minimum of $\dim_x(V)$ for all $x\in W$. If y is a point of W, we define the *codimension of W in V at y* to be the maximum attained by the codimension of $W\cap U$ in U where U is an open subset of V containing y. For our purpose, it is not important what number to be called the dimension of the empty scheme, but we shall find it convenient to agree that *the empty subscheme of any algebraic scheme can have any integer as its codimension.* For instance, we shall treat the empty subscheme as a non-singular irreducible subscheme of codimension one (or everywhere of codimension one).

In the following definitions, *X denotes a fixed non-singular irreducible algebraic scheme of dimension N.*

DEFINITION 4. *A resolution datum of type $R_I^{N,n}$ on X is an object of the following form:*

$$R_I^{N,n}=(E;\, V_1,\, V_2,\, \ldots,\, V_a;\, W),$$

where (i) *E is a reduced subscheme of X everywhere of codimension one, which has only normal crossings* (cf. Definition 3, § 1.),

(ii) *a is a non-negative integer, and V_i $(1 \leqslant i \leqslant a)$ are subschemes (not necessarily reduced, nor irreducible) of X,*

(iii) *W is a reduced subscheme (not necessarily irreducible) of X, which is non-empty and contained in the intersection of the V_i $(1 \leqslant i \leqslant a)$ in the point-set-theoretical sense, and*

(iv) $n = \dim(W)$ *and* $N = \dim(X)$.

A resolution datum of type R_{II}^N on X is an object of the following form:

$$R_{II}^N = \left(\begin{matrix} E_1, \ldots, E_b \\ c_1, \ \ldots, c_b \end{matrix} \middle| \mathfrak{J}, d \right),$$

where (i) *b and c_i $(1 \leqslant i \leqslant b)$ are non-negative integers,*

(ii) *E_i $(1 \leqslant i \leqslant b)$ are non-singular irreducible subschemes of codimension one of X, one or more of which may be the empty subscheme, $E_1 \cup \ldots \cup E_b$ has only normal crossings*(1),

(iii) *$\mathfrak{J}$ is a coherent sheaf of non-zero ideals in O_X,*

(iv) *d is a positive integer, and* $N = \dim(X)$.

DEFINITION 5. *We say that the datum*

$$R_I^{N,n} = (E;\ V_1, V_2, \ldots, V_a;\ W)$$

is resolved at a point x of W, if the following conditions are satisfied:

(i) *x is a simple point of W,*

(ii) *all the V_i $(1 \leqslant i \leqslant a)$ are normally flat along W at x, i.e. they are so within a sufficiently small open neighborhood of x in X (see Definition 1, § 1), and*

(iii) *E has only normal crossings with W at x.*

We say that the datum

$$R_{II}^N = \left(\begin{matrix} E_1, \ldots, E_b \\ c_1, \ \ldots, c_b \end{matrix} \middle| \mathfrak{J}, d \right)$$

is resolved at a point x of X if

$$\sum_{x \in E_i} c_i + \nu(\mathfrak{J}_x) < d.$$

Here the number $\nu(\mathfrak{J}_x)$ denotes the maximal non-negative integer ν such that the νth power of the maximal ideal of $O_{X,x}$ contains $\mathfrak{J}_x$.

We remark that, for a subscheme W of an algebraic scheme V, $gr_W(V)_x$ is flat over $O_{W,x}$ for a point x of W if and only if $gr_W(V)$ if flat over O_W within some neighborhood of x in W. Therefore, we have an open subset U of W such that $gr_W^n(V)_x$ is a free $O_{W,x}$-module for all non-negative integers n if and only if x is a point in U. This is the maximal open subset of W within which V is normally flat along W.

We also remark that, for a coherent sheaf of ideals $\mathfrak{J}$ on X, the set $\{x \in X : \nu(\mathfrak{J}_x) < m\}$ is an open subset of X for every integer m.

Given a resolution datum of type R_{II}^N on X as above, we set

$$S(R_{II}^N) = \{x \in X : \sum_{x \in E_i} c_i + \nu(\mathfrak{J}_x) \geqslant d\},$$

(1) The union of two (or more) subschemes is, by definition, the subscheme defined by the intersection of the ideals defining the given ones. If the given subschemes are reduced, so is the union.

$$\nu(R_{II}^N) = \max_{x \in S(R_{II}^N)} \{\nu(\mathfrak{J}_x)\},$$

$$S^*(R_{II}^N) = \{x \in S(R_{II}^N) : \nu(\mathfrak{J}_x) = \nu(R_{II}^N)\}.$$

In general, for every non-negative integer m, we set

$$S^m(R_{II}^N) = \{x \in S(R_{II}^N) : \nu(\mathfrak{J}_x) \geqslant m\}.$$

DEFINITION 6. *A resolution datum of type $R_I^{N,n}$ with closed restriction F on X is a pair*

$$(R_I^{N,n}, F)$$

of a resolution datum of type $R_I^{N,n}$ on X, described above, and a non-singular irreducible subscheme F of codimension one of X, such that

(*i*) *$E \cup F$ has only normal crossings, and*

(*ii*) *F does not contain any irreducible component of W.*

A resolution datum of type $R_I^{N,n}$ with open restriction U on X is a pair

$$(R_I^{N,n}, U)$$

of a resolution datum of type $R_I^{N,n}$ on X, described above, and an open subset U of W such that

(*i*) *U has non-empty intersection with every irreducible component of W, and*

(*ii*) *the datum $R_I^{N,n}$ is resolved at every point of U.*

A resolution datum of type R_{II}^N with closed restriction F on X is a pair

$$(R_{II}^N, F)$$

of a resolution datum of type R_{II}^N on X, described above, and a non-singular irreducible subscheme F of codimension one of X such that $E_1 \cup \ldots \cup E_b \cup F$ has only normal crossings.

We shall also consider a resolution datum of type R_{II}^N *without restriction* on X.

Let B be a subscheme of a reduced scheme Y, nowhere dense in Y, and $\mathfrak{J}$ the sheaf of ideals of B on Y, i.e. O_B is canonically isomorphic to $O_Y/\mathfrak{J}$. Then there exists a unique *reduced modification* $f: Y' \to Y$ (i.e. a morphism which induces an isomorphism of a dense open subset of Y' onto such an open subset of Y, where Y' is a reduced scheme) which satisfies the universal mapping property, (*a*) and (*b*) of Example IV, § 1, for reduced modifications of Y. This reduced modification, determined by B, will be called *the monoidal transformation of Y with center B*. If Y is an algebraic scheme, then so is Y'. We are interested in the case where Y is a non-singular algebraic scheme and B is a non-singular irreducible subscheme of Y, in which case Y' is necessarily a non-singular algebraic scheme. In the sequel, it is important that a monoidal transformation is always understood as a reduced modification *together with a specific center*, so that the monoidal transformations of a scheme Y is in a one-to-one correspondence with the nowhere dense subschemes of Y which are called centers of transformations.

Let X be as above, and B any non-singular irreducible subscheme of X

which is different from X. Let V be any subscheme of X. If $f_B: X' \to X$ denotes the monoidal transformation of X with center B (as well as the morphism itself), then *the (total) transform* $f_B^{-1}(V)$ is the subscheme of X' defined by the sheaf of ideals $f_B^{-1}(\mathfrak{J})$ on X' where $\mathfrak{J}$ denotes the sheaf of ideals of V on X. We also define and denote by $f_B^{-1}(V)_0$ *the restricted transform of V by f_B* as follows: The morphism f_B induces an isomorphism $\bar{f}: X' - f_B^{-1}(B) \to X - B$. Then $\bar{f}^{-1}(V-B)$ is a well-defined subscheme of $X' - f_B^{-1}(B)$, and there exists *the smallest* subscheme V' of X' such that $\bar{f}^{-1}(V-B)$ is the restriction of V' to $X' - f_B^{-1}(B)$. (We say that a subscheme is *smaller* than another if the defining sheaf of ideals of the first contains that of the second.) This subscheme, uniquely determined by V and f_B (i.e. V and B), V' of X' is the restricted transform of V by the monoidal transformation f_B.

DEFINITION 7. *We say that a monoidal transformation f_B of X with center B is permissible for the resolution datum $R_I^{N,n}$ on X, described above, if B is non-singular and irreducible and, at the same time, the object of the form*

$$(E;\, V_1, V_2, \ldots, V_a, W;\, B)$$

is a resolution datum of type $R_I^{N,n}$ on X which is resolved everywhere (i.e. at every point of B).

We say that f_B is permissible for the resolution datum $R_I^{N,n}$ with closed (or open) restriction F (or U) on X, described above, if f_B is permissible for $R_I^{N,n}$ and, at the same time, B is contained in F (or in $W-U$).

We say that f_B is permissible for the resolution datum R_{II}^N (without restriction) on X, described above, if B is a non-singular irreducible subscheme of X contained in $S^*(R_{II}^N)$ and, at the same time, $E_1 \cup \ldots \cup E_b$ has only normal crossings with B.

We say that f_B is permissible for the resolution datum R_{II}^N with closed restriction F on X, described above, if f_B is permissible for R_{II}^N and, at the same time, B is contained in and different from F.

DEFINITION 8. *Let R be a resolution datum on X, either $(R_I^{N,n}, F)$ or $(R_I^{N,n}, U)$ or (R_{II}^N, F) or R_{II}^N, where*

$$R_I^{N,n} = (E;\, V_1, V_2, \ldots, V_a;\, W)$$

and

$$R_{II}^N = \left(\begin{matrix} E_1, \ldots, E_b \\ c_1, \ldots, c_b \end{matrix} \middle| \, \mathfrak{J}, d \right).$$

(See Definitions 4 and 6.) Let f_B be a monoidal transformation of X with center B which is permissible for the resolution datum R, say $f_B: X' \to X$. Then we set: $E' = f_B^{-1}(E)_0$, $V_i' = f_B^{-1}(V_i)_0$ $(1 \leqslant i \leqslant a)$, $W' = f_B^{-1}(W)_0$, $E_i' = f_B^{-1}(E_i)_0$ $(1 \leqslant i \leqslant b)$ and $E_{b+1}' = f_B^{-1}(B)$. Let $F' = f_B^{-1}(F)_0$ and $U' = f_B^{-1}(U) = f_B^{-1}(U) \cap W'$. We define the transform $f_B^*(R)$ of the resolution datum R on X by the permissible monoidal transformation f_B as follows:

$$f_B^*(R_I^{N,n}, F) = (f_B^*(R_I^{N,n}), F'),$$

$$f_B^*(R_I^{N,n}, U) = (f_B^*(R_I^{N,n}), U'),$$

where

$$f_B^*(R_I^{N,n}) = (E' \cup E_{b+1}';\, V_1', V_2', \ldots, V_a';\, W'),$$

$$f_B^*(R_{II}^N, F) = (f_B^*(R_{II}^N), F'),$$

where $$f_B^*(R_{II}^N) = \left(\begin{matrix} E_1', \ldots, E_b', E_{b+1}' \\ c_1, \ldots, c_b, \ c_{b+1} \end{matrix} \middle| f_B^{-1}(\mathfrak{J})\, \mathfrak{P}^{-m}, d\right)$$

with $m = \nu(R_{II}^N)$, $c_{b+1} = \sum_{B \subseteq E_i} c_i + m - d$, *and* $\mathfrak{P}$ = *the sheaf of ideals of* E_{b+1}' *on* X'.

We remark that the transform of any resolution datum on X by a permissible monoidal transformation $f_B : X' \to X$ is again a resolution datum of the same type on X'.

Now that we have defined the permissibility of monoidal transformation and the transform of datum by such a transformation, we can speak of a finite succession of permissible monoidal transformations. More precisely as follows:

Let R be a resolution datum on X, and

$$f = \{f_i : X_i \to X_{i-1} \quad (1 \leqslant i \leqslant t, X_0 = X)\}$$

a finite succession of monoidal transformations f_i with centers B_{i-1} in X_{i-1}. We say that *the succession f is permissible for R if there exists a resolution datum R_i on X_i for $0 \leqslant i \leqslant t$ such that*

(1) *f_i is permissible for R_{i-1} $(1 \leqslant i \leqslant t)$, and*

(2) *$R_0 = R$ and $R_i = f_i^*(R_{i-1})$ $(1 \leqslant i \leqslant t)$.*

We shall use the same symbol f for the succession of monoidal transformations as well as for the composed morphism $X_t \to X$, whenever the context makes the sense clear. We shall call the above resolution datum R_t on X_t *the transform of R by the permissible succession f for R*, and denote it by $f^*(R)$. It should be noted that the transform $f^*(R)$ is determined by the succession f but not by the composed morphism. Namely, the way of factoring the morphism into monoidal transformations with specified centers is essential in the definition of $f^*(R)$.

The following are the fundamental resolution theorems of this work.

THEOREM $I_F^{N,n}$. (*The separation theorem of the first type.*)

Given a resolution datum $(R_I^{N,n}, F)$ of type $R_I^{N,n}$ with closed restriction F on a non-singular irreducible algebraic scheme X, there exists a permissible succession of monoidal transformations, $f : X' \to X$, of X for $(R_I^{N,n}, F)$ such that, if

$$f^*(R_I^{N,n}, F) = (f^*(R_I^{N,n}), F')$$

with $$f^*(R_I^{N,n}) = (E', V_1', V_2, \ldots, V_a'; W'),$$

then W' and F' have no common points.

THEOREM $I_U^{N,n}$. (*The resolution theorem of the first type.*)

Given a resolution datum $(R_I^{N,n}, U)$ of type $R_I^{N,n}$ with open restiction U on a non-singular irreducible algebraic scheme X, there exists a permissible succession of monoidal transformations, $f : X' \to X$, of X for $(R_I^{N,n}\, U)$ such that $f^*(R_I^{N,n})$ is resolved everywhere.

THEOREM II_F^N. (*The separation theorem of the second type.*)

Given a resolution datum (R_{II}^N, F) of type R_{II}^N with closed restriction F on a non-singular irreducible algebraic scheme X, such that $m = \nu(R_{II}^N) > 0$ and $F \nsubseteq S^*(R_{II}^N) = S^m(R_{II}^N)$, there exists a permissible succession of monoidal transformations, $f : X' \to X$, of X for (R_{II}^N, F) such that, if

$$f^*(R_{II}^N, F) = (f^*(R_{II}^N), F'),$$

then $S^m(f^*(R_{II}^N))$ *and* F' *have no common points.*

THEOREM II^N. (*The resolution theorem of the second type.*)

Given a resolution datum R_{II}^N *of type* R_{II}^N *without restriction on a non-singular irreducible algebraic scheme* X, *there exists a permissible succession of monoidal transformations,* $f: X' \to X$, *of* X *for* R_{II}^N *such that* $f^*(R_{II}^N)$ *is resolved everywhere.*

In the above four theorems, N and n are arbitrary integers such that $N > n \geqslant 0$. We first see that Theorems $I_F^{N,0}$, $I_U^{N,0}$ and II_F^1 are trivially true, and that Theorem II^1 can be easily verified. The proof of the four theorems for arbitrary N and n can be done by showing the following basic implications:

(*A*) Given (N, n) such that $N > n \geqslant 1$,

$$\text{Theorems} \quad \left\{\begin{array}{lll} I_F^{N',n'} & \text{with} & n' < N' < N, \\ I_U^{N,n''} & \text{with} & n'' < n, \\ II^{N'} & \text{with} & N' < N \end{array}\right\} \Rightarrow \text{Theorem } I_F^{N,n},$$

(*B*) Given (N, n) such that $N > n \geqslant 1$,

$$\text{Theorems} \quad \left\{\begin{array}{lll} I_U^{N',n'} & \text{with} & n' < N' < N, \\ I_U^{N,n''} & \text{with} & n'' < n, \\ I_F^{N'',n''} & \text{with} & \left\{\begin{array}{l} N'' \leqslant N, \\ n'' \leqslant n, \\ n'' < N'', \end{array}\right. \\ II^{N'} & \text{with} & N' < N \end{array}\right\} \Rightarrow \text{Theorem } I_U^{N,n},$$

(*C*) Given $N \geqslant 2$,

$$\text{Theorems} \quad \left\{\begin{array}{lll} II_F^{N'} & \text{with} & N' < N, \\ I_U^{N'',n''} & \text{with} & n'' < N'' \leqslant N, \\ II^{N'} & \text{with} & N' < N \end{array}\right\} \Rightarrow \text{Theorem } II_F^N,$$

(*D*) Given $N \geqslant 2$,

$$\text{Theorems} \quad \left\{\begin{array}{lll} II^{N'} & \text{with} & N' < N, \\ I_U^{N'',n''} & \text{with} & n'' < N'' \leqslant N, \\ II_F^{N''} & \text{with} & N'' \leqslant N \end{array}\right\} \Rightarrow \text{Theorem } II^N.$$

The proof of these four implications requires certain non-trivial analysis of singularities. We entirely omit this analysis here, and those readers who want to find the detailed proof should refer to either the writer's mimeographed seminar note or the forthcoming papers in *Annals of Mathematics*. We simply point out that our proof does not depend upon any preexisting non-trivial results, such as the Local Uniformizing Theorem of Zariski, and hope that the diagram of implications, shown above, indicates at least the close relations among various aspects of the resolution problems. The three Theorems, stated and paragraphed in § 1, can be deduced from the four Theorems in this section by standard GAGA techniques and com-

paratively simple inductions, and we hope that they may serve to clarify some geometric significances which underlie the above inductive formulation of resolution problems.

REFERENCES

[1]. Levi, B., Sulla riduzione delle singolarità puntuali delle superficie algebriche dello spazio ordinario per transformazioni quadratiche. *Ann. Mat. Pura Appl. II*, 26 (1897).

[2]. —— Risoluzione delle singolarità puntuali delle superficie algebriche. *Atti Accad. Sci. Torino*, 33 (1897).

[3]. Zariski, O., The reduction of the singularities of an algebraic surface, *Ann. Math.*, 40 (1939).

[4]. —— Algebraic surfaces. *Ergebnisse d. Math.*, III, 5). Berlin, 1935.

[5]. —— Reduction of singularities of algebraic three-dimensional varieties. *Ann. Math.*, 45 (1944).

[6]. —— Local uniformization on algebraic varieties. *Ann. Math.*, 41 (1940).

[7]. Grothendieck, A., Eléments de géométrie algébrique. I. *Publ. Math. (I.H.E.S.) Paris*, 4 (1960).

STRUCTURE THEOREMS OF MODULAR VARIETIES

By JUN-ICHI IGUSA

If we try to generalize various results on elliptic modular functions, we will have to know bounded domains (and also convenient unbounded domains isomorphic to bounded domains) which can be considered as higher dimensional analogues of the interior of the unit circle. Since otherwise we can do nothing, we assume that the bounded domains are homogeneous. Then, to get a reasonable theory (especially to have quotient varieties of finite invariant volumes by properly discontinuous groups), we are forced to assume that the bounded homogeneous domains are symmetric in the sense of E. Cartan. As a corollary of his great work on the theory of symmetric Riemann spaces, we have a complete list of such domains. Namely we know the so-called four main types and two exceptional types of irreducible bounded symmetric domains. Among these domains, one type (type IV of E. Cartan and type III of Siegel [5]) is particularly interesting. The pth representative ($p=1,2,\ldots$) of this type is an open set $\mathfrak{D}$ in $\mathbf{C}^n$ ($n=p(p+1)/2$) whose points are p-by-p complex symmetric matrices z such that $1_p-{}^t\bar{z}z$ are positive-definite. This domain is isomorphic to the Siegel upper-half plane of degree p, i.e. an unbounded domain $\mathfrak{S}$ in $\mathbf{C}^n$ whose points are p-by-p complex symmetric matrices z with positive-definite imaginary parts. The very reason why we are so attached to this type is that $\mathfrak{S}$ is the space of deformations of abelian varieties with certain polarizations; and, in the case $p=1$, i.e. in the theory of elliptic modular functions, this fact is fundamental. Furthermore, in the full group of automorphisms of $\mathfrak{S}$, which is $\mathrm{Sp}(p,\mathbf{R})$ (more precisely $\mathrm{Sp}(p,\mathbf{R})$ modulo its center $\pm 1_{2p}$) acting on $\mathfrak{S}$ as

$$z\to(az+b)\,(cz+d)^{-1},$$

the discrete group $\Gamma=\mathrm{Sp}(p,\mathbf{Z})$ called the Siegel modular group of degree p (and of level 1) is particularly interesting. The reason is that the corresponding quotient variety $\Gamma\backslash\mathfrak{S}$ is precisely the variety of p-dimensional abelian varieties with polarizations similar to canonical polarizations of jacobian varieties. The structure of the variety $\Gamma\backslash\mathfrak{S}$ is therefore of considerable interest to algebraic geometers. We ask ourselves now what we mean by saying that we know the structure of a complex variety. If it is an affine variety, we can certainly say that we know the structure of the variety once we know its co-ordinate ring. Similarly, if the variety is embeddable in a projective space as its closed subvariety, we can say that we know the structure of the variety once we know the co-ordinate ring of a projective embedding by a (preferably intrinsic) polarization. Coming back to the variety $\Gamma\backslash\mathfrak{S}$, although it is neither affine (except for $p=1$) nor compact, it has been proved by a combined effort of several mathematicians that it has a rather natural compactification to a normal projective variety [1]. Furthermore, the projective embedding is given by Siegel modular forms of

a sufficiently high weight w (which we assume to be even), i.e. by holomorphic functions ψ on $\mathfrak{S}$ which satisfy the functional equation

$$\psi((az+b)\,(cz+d)^{-1})=\det(cz+d)^{w}\psi(z)$$

for every operation of Γ (and which satisfy an additional condition at infinity when $p=1$). Therefore the structure of $\Gamma\backslash\mathfrak{S}$ will certainly be known once we know the structure of the graded ring S generated by the Siegel modular forms. The importance of S partly comes also from the fact that meromorphic functions on $\Gamma\backslash\mathfrak{S}$ (on the compactification of $\Gamma\backslash\mathfrak{S}$ when $p=1$) are Siegel modular functions, i.e. quotients of homogeneous elements of S of the same degree. Now the aforesaid results imply that S is a positively graded normal integral ring of finite type over $S_0=\mathbf{C}$. Beyond this fact, we know very little about the ring S. In the special case $p=1$, it is a classical theorem that S is generated by Eisenstein series of weights 4 and 6. In the case $p=2$, it was proved by Witt and Maass that $S_2=0$ and that S_4, S_6, S_8 are 1-dimensional. Also it was known more or less that the field of Siegel modular functions is a purely transcendental extension of $\mathbf{C}$. In fact, this can be proved by identifying the field of modular functions with the field of absolute invariants of binary sextics.

We shall now explain our contributions to this subject. We investigated first the invariants of binary sextics in connection with moduli of curves of genus two. This was done in a most satisfactory manner. Namely we determined the structure of the variety of moduli as a scheme over Spec($\mathbf{Z}$) [2]. In particular, its complex fibre is the quotient variety of $\mathbf{C}^3$ by a cyclic group of order 5 acting as

$$(z_1,z_2,z_3)\to(\zeta z_1,\zeta^2 z_2,\zeta^3 z_3)\quad(\zeta^5=1).$$

We then applied this result to determining the structure of $\Gamma\backslash\mathfrak{S}$ (in the case $p=2$) [3]. It turned out that $\Gamma\backslash\mathfrak{S}$ has a covering by the variety of moduli and by another affine variety with a similar structure. This information enabled us to determine S and it was shown that S is generated by Eisenstein series

$$\psi_w(z)=\sum\det(cz+d)^{-w}$$

of weights $w=4, 6, 10, 12$. Consequently the compactification of $\Gamma\backslash\mathfrak{S}$ is the quotient variety of $\mathbf{C}^4-0$ by the multiplicative group $\mathbf{C}^*$ acting as

$$(t_1,t_2,t_3,t_4)\to(\varrho^2 t_1,\ \varrho^3 t_2,\ \varrho^5 t_3,\ \varrho^6 t_4)\quad(\varrho\in\mathbf{C}^*).$$

The aforesaid affine varieties are the varieties respectively defined by $t_3\neq 0$ and $t_4\neq 0$. The underlying topological space of this type of varieties is a quotient space of a sphere by a circle and the additive structure of the integer cohomology ring was shown by Floyd and Conner to be the same as that of a (complex) projective space of the same dimension. The multiplicative structure was recently determined by Tramer and in the present case, if g_i are the generators in dimensions i for $i=2,4,6$, the cup-product formulas are

$$g_2\cup g_2=5g_4,\quad g_2\cup g_4=30g_6.$$

We believe that the structure of the variety $\Gamma\backslash\mathfrak{S}$ is now sufficiently known to investigate various things, e.g. the so-called modular correspondences and their fixed-varieties etc.

Going back to the variety of moduli of curves of genus two, we shall outline (for those people who are interested only in the complex case) a shorter proof of the structure theorem. We observe that the generating function of the graded ring of projective invariants can be determined easily by a computation of characters of $SL(2, \mathbf{C})$ and we get

$$\sum_{m=0}^{\infty} (\text{number of invariants of degree } m)\, t^m$$
$$= (1+t^{15})((1-t^2)(1-t^4)(1-t^6)(1-t^{10}))^{-1}.$$

The invariants A, B, C, D of degrees 2, 4, 6, 10 generate a subring of the graded ring of even projective invariants and the above formula shows that the subring is actually the ring of all even invariants. We then recall a simple lemma which can be stated as follows. Suppose that k is an algebraically closed field and G an algebraic group over k such that every k-representation is completely reducible. Then, for every affine k-algebra $R = k[x_1, \ldots, x_n]$ on which G acts linearly (in x_i), the natural morphism $\mathrm{Spec}(R) \to \mathrm{Spec}(R^G)$ is surjective. We can apply this lemma (taking $k = \mathbf{C}, G = SL(2, \mathbf{C})$) and the previous remark to show that, if we evaluate A, B, C, D at the sextic

$$xy(x-y)\,(x-\lambda_1 y)\,(x-\lambda_2 y)\,(x-\lambda_3 y),$$

then λ_1, λ_2, λ_3 are integral over the subring of elements of degree 0 in the ring of fractions of $\mathbf{C}[A, B, C, D]$ with respect to powers of D. The structure theorem of the complex variety of moduli is an immediate consequence of this fact.

Finally we add a few words about the problem of determining the structure of S in general. We know that we can restrict modular forms in $\mathfrak{S}$ to its boundary components and in this way we get homomorphisms of S from degree p to smaller degrees. The fact is that these homomorphisms known as Siegel operators are almost surjective. Therefore, if we know the kernel of the Siegel operator from degree p to degree $p-1$, we can roughly determine the structure of S by an induction on p. Now if we restrict to additive structures, we have the so-called Selberg trace formula which expresses the dimension of each homogeneous part of the kernel by a certain integral extended over the fundamental domain of Γ [4]. It might be that such integrals are connected with some group characters and with something much simpler in non-archimedean geometry. Even if this is just wishful thinking, still we should try seriously to calculate the trace formula explicitly. In fact, it is almost certain that the multiplicative structure of S is determined at most up to a finite number of possibilities by its additive structure.

REFERENCES

[1]. *Séminarie* H. CARTAN, Fonctions automorphes. *E.N.S.*, 1957–58.

[2]. IGUSA, J., Arithmetic variety of moduli for genus two. *Ann. Math.*, 72 (1960), 612–649.

[3]. —— On Siegel modular forms of genus two. *Amer. J. Math.*, 84 (1962), 175–200.

[4]. SELBERG, A., Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Ind. Math. Soc.*, 20 (1956), 47–87.

[5]. SIEGEL, C. L., Analytic functions of several complex variables. *I.A.S.*, 1948–49.

PROJECTIVE INVARIANTS OF PROJECTIVE STRUCTURES AND APPLICATIONS

By DAVID MUMFORD

The basic problem that I wish to discuss is this: if V is a variety, or scheme, parametrizing the set, or functor, of all structures of some type in projective n-space $\mathbf{P}_n$, then the group $PGL(n)$ of automorphisms of $\mathbf{P}_n$ acts on V. Then under what conditions does there exist a quotient or orbit space $V/PGL(n)$, i.e. when can we construct enough "projective invariants" for these structures? For example, let V parametrize the set of hypersurfaces of degree m, with certain types of singularities; or let V parametrize the set of tri-canonical space curves of given genus, or even n-canonical surfaces with at most "negligible singularities" [1]; or let V parametrize the set of 0-cycles of degree m in $\mathbf{P}_n$; or let V parametrize the set of all morphisms of a fixed scheme X into $\mathbf{P}_n$. Moreover, I wish to illustrate how such questions are one essential step in several basic existence and construction problems of algebraic geometry.

One approach to this problem is afforded by the invariant theory of the representations of reductive groups. Here you generalize the problem first: consider an arbitrary action of an algebraic group G on a variety (or scheme) V and seek an orbit space V/G. Then you specialize the problem by (a) assuming G is reductive, (b) restricting attention to quasi-projective orbit spaces V/G. Of course, if, in particular, the characteristic is 0, $PGL(n)$ is reductive. Now suppose you have a projective embedding $V \subset \mathbf{P}_N$. If V is a normal variety, and if this embedding is defined by a complete linear system on V, it is possible to prove that the action of G on V extends to an action of G on $\mathbf{P}_N$. In any case, if this occurs, I say that G *acts linearly* on $V \subset \mathbf{P}_N$.

Now, if the action of G is linear, its action is induced by a linear and unimodular representation of some finite covering G^* of G on the affine cone $\mathbf{A}^{N+1}$ over the ambient $\mathbf{P}_N$. Then I make the definition:

A point $x \in V$ is *stable* for the action of G on V, relative to the embedding $V \subset \mathbf{P}_N$, if for one (and hence all) homogeneous points $x^* \in \mathbf{A}^{N+1}$ over x, (i) the stabilizer of x^* is a finite subgroup of G^*, and (ii) the orbit of x^* under G^* is closed in $\mathbf{A}^{N+1}$.

Now assume that a reductive algebraic group G acts linearly on $V \subset \mathbf{P}_N$. The fundamental theorem is this:

THEOREM. *The set of stable points forms an open set U in V, and a quasi-projective orbit space U/G exists.*

In case G is semi-simple and of characteristic 0, I can say more. First, let us call the action $\alpha: G \times V \to V$ of G on V a *proper action* if the morphism $\alpha \times Pr_2: G \times V \to V \times V$ is proper (see [2]). Then:

THEOREM. (1) *A point $x \in V$ is stable if and only if there is a hypersurface section H of V which is invariant under G, and such that $x \notin H$ and the orbit of x is closed in $V - H$.*

(2) *The action of G on the set of stable points U is proper.*

(3) *If G acts properly on V, and if a quasi-projective orbit space V/G exists, then for some projective embedding $V \subset \mathbf{P}_N$, every point of V is stable.*

Finally, if G is any reductive algebraic group in characteristic O, then I can analyze the manner in which stability breaks down in the following way:

THEOREM. *If $x \in V$ is not stable for the action of G, then there is a Borel subgroup $B \subset G$ and a 1-parameter subgroup $\mathbf{G}_m \subset B$ such that, if $\mathbf{G}_m$ is any conjugate of $\mathbf{G}_m$ in B, then x is not stable for the action of $\mathbf{G}_m$ on V.*

These results are not definitive: I conjecture that they all are valid for semi-simple groups of any characteristic. However, they suggest the kind of answer that should be found for the original question: to every type of structure in $\mathbf{P}_n$, there is a stability condition which is sufficient and, in general, necessary for the existence of enough projective invariants to classify these structures; and, moreover, this stability condition is always of the form: there is no flag in $\mathbf{P}_n$ which has "too high an order of contact" with the given structure.

The following case has been worked out exhaustively by myself and J. Tate in all characteristics, and even over the scheme of integers: let $V = (\mathbf{P}_n)^m$, i.e. V parametrizes ordered 0-cycles of degree m in $\mathbf{P}_n$. Then relative to the Segre embedding of V, a point $x = (x_1, x_2, \ldots, x_m)$ is stable if and only if:

For all linear subspaces $\mathrm{L} \subset \mathbf{P}_n$, then:

$$\frac{(\text{number of points } x_i \text{ in subspace } L)}{(\text{total number of points } x_i)} < \frac{\dim (L) + 1}{n+1}.$$

Then in all characteristics, or even over the scheme of integers, the set of stable 0-cycles forms an open set $U \subset (\mathbf{P}_n)^m$, and a quasi-projective orbit space $U/PGL(n)$ exists. In fact, *U is a principle fibre bundle over $U/PGL(n)$.* Our techniques are entirely elementary (see [3] and [6]).

I would like to illustrate how this simple result can be used to prove the existence of (i) a moduli scheme for polarized abelian varieties, and (ii) (according to a suggestion of Grothendieck) the Picard scheme of any variety.(1) First, let me fix some notations: if $U \subset (\mathbf{P}_n)^m$ is as above, let $Q_{n,m} = U/PGL(n)$. Second, since U is a principal fibre bundle over $Q_{n,m}$ with group $PGL(n)$, if $PGL(n)$ acts on any scheme X, we can form an associated fibre bundle with fibre X; in particular, if $X = \mathbf{P}_n$, denote the associated fibre bundle by $P_{n,m}$. Let $\pi: P_{n,m} \to Q_{n,m}$ be the bundle morphism. It is easy to see that the bundle $P_{n,m}/Q_{n,m}$ has m distinguished sections $s_i: Q_{n,m} \to P_{n,m}$ associated to the maps which take the m-tuple $(x_1, x_2, \ldots, x_m)$ to its ith factor x_i. Intuitively, regard $P_{n,m}$ plus the collection of sections (s_i) as the universal family of projective n-spaces with given stable 0-cycle of degree m.

To simplify the treatment of the moduli problem, let us consider only the question of finding a quasi-projective variety $M_{g,d}$ whose points parametrize "naturally" the set of all abelian varieties A of dimension g, plus a very ample(2) divisor class D such that $(D^g) = d \cdot g!$ For the question of

(1) In fact, it can be used to prove much stronger results on the existence of relative Picard schemes: see [4].

(2) in the sense of Grothendieck, i.e. induced via a projective embedding.

making explicit the sense of the word "naturally", and the question of replacing D by its numerical equivalence class, see [6]. Now pick any n such that the characteristic does not divide n, and $n > d \cdot \sqrt{g!}$. Let $\nu = n^{2g}$. Then consider the following set:

Subvarieties $A \subset P_{d-1,\nu}$ such that

(*a*) $\pi(A)$ is a single point $q \in Q_{d-1,\nu}$,
(*b*) the degree of A in $\pi^{-1}(q)$ equals $d \cdot g!$, and the dimension of A is g,
(*c*) A admits a structure of abelian variety such that $\{s_i(q)\}$, for $i = 1, 2, \ldots, \nu$ is the set of points of order n.

Then, on the one hand, this set is parametrized naturally by a locally closed subset of a suitable Chow variety. But, on the other hand, it is isomorphic to the set of abelian varieties A of dimension g, plus a very ample divisor class D such that $(D^g) = d \cdot g!$, *plus* an ordering of the set of points of order n. [Namely, one can prove that the inequality $n > d \cdot \sqrt{g!}$ insures that the set of points of order n form a stable 0-cycle in $\mathbf{P}_{d-1}$ via the embedding defined by the complete linear system $|D|$. Let q be the equivalence class of this 0-cycle mod $PGL(d-1)$; then there is a unique identification of $\pi^{-1}(q)$ with the $\mathbf{P}_{d-1}$ ambient to A, under which $s_i(q)$ corresponds to the ith point of order n. Then A is embedded in $\pi^{-1}(q) \subset P_{d-1,\nu}$ via this identification.] Therefore, the set of all A, plus D alone, is parametrized by the quotient of this set by the group of permutations of ν letters. As such a quotient of a quasi-projective variety is well-known to exist, this solves the problem.

To simplify the treatment of the Picard scheme of a projective variety V, let us consider only the construction of the reduced and connected Picard scheme. The question then may be loosely described as that of finding a variety P whose points parametrize "naturally" the set of all Cartier divisor classes D on V which are algebraically equivalent to 0. But this is equivalent to doing the same for the set of all Cartier divisor classes D on V which are algebraically equivalent to some fixed very ample divisor D_0. But if D_0 is chosen sufficiently ample, and if we associate to each divisor class D the morphism $\phi: V \to \mathbf{P}_n$ defined by the complete linear system $|D|$, then this set will be isomorphic to the set of orbits under the group $PGL(n)$ in the set of morphisms $\phi: V \to \mathbf{P}_n$ which are algebraically equivalent to a fixed $\phi_0: V \to \mathbf{P}_n$. This is a problem of the type originally posed.

It can be reduced to the 0-cycle problem already solved as follows: pick a large number of sufficiently generic points $x_1, x_2, \ldots x_N$ in V, i.e. such that:

For all $\phi: V \to \mathbf{P}_n$ algebraically equivalent to ϕ_0, the set of points ϕx_1, $\phi x_2, \ldots, \phi x_N$ is a stable 0-cycle in $\mathbf{P}_n$.

Then consider the following set:

Morphisms $\phi: V \to P_{n,N}$ such that

(*a*) $\pi \circ \phi$ maps V to a single point $q \in Q_{n,N}$,
(*b*) $s_i(q) = \phi(x_i)$,
(*c*) ϕ, as a morphism from V to $\pi^{-1}(q) \cong \mathbf{P}_n$ is algebraically equivalent to ϕ_0.

Then on the one hand this set is parametrized by a locally closed subset of a suitable Chow variety (i.e. via the graph of ϕ), and on the other hand it is isomorphic to the set of orbits described above.

Here is another example of a stability condition and the resulting quotient theorem. Assume the characteristic is 0, and consider, instead of sequences of points in $\mathbf{P}_n$, sequences of linear subspaces of any dimension. Thus, if $Grass_{k,n}$ stands for the Grassmannian of k-dimensional linear subspaces of $\mathbf{P}_n$, I ask for orbit spaces of the type $(Grass_{k,n})^m/PGL(n)$. Then, in fact, relative to the usual Plücker embedding of the Grassmannian, and to the Segre embedding of this product, it turns out that a point $x=(L_1,L_2,\ldots,L_m)$ of $(Grass_{k,n})^m$ is stable if and only if:

For all linear subspaces $L\subset P_n$, then:

$$\frac{\sum_i(\dim\,(L\cap L_i)+1)}{\sum_i(\dim\,L_i+1)}<\frac{(\dim\,L+1)}{(n+1)}.$$

Then, by the fundamental theorem, the set of stable points forms an open set U, and an orbit space $U/PGL(n)$ exists.

This result can be applied to the problem of classifying vector bundles over a variety in exactly the same way as the result on 0-cycles was applied to the problem of classifying line bundles, i.e. Cartier divisor classes. Of course, it is well-known that the set of vector bundles even over an algebraic curve is not a separated space; in fact, it is not even locally separated, because of the "jump" phenomenon noted by Kodaira and Spencer [5]. However, again, a basic stability condition avoids all these difficulties. For simplicity, let us consider only vector bundles over a fixed-curve C.

DEFINITION. A vector bundle E is *stable* if for all sub-bundles F,

$$\operatorname{Deg} c_1(F)<\operatorname{Deg} c_1(\mathrm{E})\cdot\frac{\operatorname{rank} F}{\operatorname{rank} E},$$

where c_1 denotes the first chern class.

In other words, a vector bundle is stable if all its subbundles are "less ample" than itself. To illustrate the stability condition, let me mention its simplest properties:

(i) If L is a line bundle, then E is stable if and only if $E\otimes L$ is stable; moreover, E is stable if and only if $\check{E}$ is stable.
(ii) If E_1 and E_2 are two vector bundles, $E_1\oplus E_2$ is never stable.
(iii) A line bundle is always stable.
(iv) If a vector bundle E of rank 2 is not stable, then either E is isomorphic to $L_1\oplus L_2$, *or* there is a unique sub-bundle L for which $\geqslant$ holds in the definition and E can be canonically described as an extension.

Then I can prove the following theorem:

THEOREM. *The set of all stable vector bundles of rank* r *over a fixed curve* C *in characteristic* 0 *is "naturally" isomorphic to the set of points of a non-singular quasi-projective variety* $V_r(\mathrm{C})$.

A more complicated example of a stability condition is given by the action of $PGL(2)$ on the variety of plane curves of degree n. There seems to be no simple general rule describing when a plane curve is stable; however, I can prove that if $n\geqslant 3$, then at least every non-singular curve is stable. For low values of n, the precise answer is given by:

n	Stable curves
1, 2	None.
3	Non-singular.
4	No triple points and no tacnodes.
5	No triple points with 3 coincident tangents, or with 2 tangents forming a tacnode.

To conclude, I want to pose a question that seems to me to be the most interesting problem in extending the results discussed above. Moreover, I think this problem is the central one on the road to solving the general problem of algebraic moduli of polarized non-singular varieties. The question is: *when is the Chow point of a non-singular subvariety* $V \subset \mathbf{P}_n$ *stable* (*in the usual projective embedding of the Chow variety*)? Perhaps more reasonable is the stabilized form of this problem: Given a subvariety $V \subset \mathbf{P}_n$, when is there an n_0 such that if $n \geqslant n_0$, the Chow point of V in the n-ple embedding $V \subset \mathbf{P}_N$ is stable? I have no conjecture to make.

REFERENCES

[1]. Artin, M., Some numerical criteria for the contractability of curves on an algebraic surface. *Amer. J. Math.*, 84 (1962), 485.

[2]. Palais, R., On the existence of slices for actions of non-compact Lie groups. *Ann. Math.*, 73 (1961), 295.

[3]. Mumford, D., An elementary theorem in geometric invariant theory. *Bull. Amer. Math. Soc.*, 67 (1961), 483.

[4]. Grothendieck, A., *Séminaire Bourbaki*, exp. 232 and 236, 1962.

[5]. Kodaira, K. & Spencer, D. C., On deformations of complex analytic structures. *Ann. Math.*, 67 (1958), 328.

Most of the material discussed above will be published in:

[6]. Mumford, D., Geometric invariant theory. (Forthcoming.)

TRANSFORMATION OF MARKOFF PROCESSES

By G. A. HUNT

In developing the theory of Martin boundaries for diffusion processes or for Markoff processes with countably many states it has sufficed to treat in detail the exit boundary, deriving by duality the properties of the entrance boundary; this state of affairs apparently persists so long as one discusses Markoff processes for which the exit boundary has a straightforward definition. On the other hand, the construction of the entrance boundary makes sense in situations where one is puzzled how to define the exit boundary, so that finally the entrance boundary must be dealt with directly. This can be done rather simply, as I shall describe, using ideas from recent papers by Dynkin, Hunt and Ray (see the references).

Let $P_t(x, dy)$ be a transition function on the separable locally compact space $\mathcal{E}$. To lighten the exposition I shall assume that whenever f is a continuous function on $\mathcal{E}$ vanishing at infinity the function $P_t f$,

$$P_t f(x) = \int_{\mathcal{E}} P_t(x, dy) f(y), \tag{1}$$

has the same properties and tends uniformly to f as t decreases to zero. In order to define the Martin boundary one must also impose some condition on the kernel $U_x(dy)$,

$$U_x(E) = \int_0^\infty P_t(x, E)\, dt \qquad (E \subset \mathcal{E}) \tag{2}$$

I shall assume that whenever E is compact $U_x(E)$ is bounded in x.

A positive measure η on $\mathcal{E}$ is excessive if it is finite on compact sets and if for each t it dominates the measure ηP_t,

$$\eta P_t(E) = \int_{\mathcal{E}} \eta(dx)\, P_t(x, E). \tag{3}$$

The nontrivial excessive measures form a cone $\mathcal{M}$ in the space of measures on $\mathcal{E}$, which we suppose given the weak topology. For our purposes one should identify proportional excessive measures; so we take $\mathcal{M}^*$ to be the space of rays of $\mathcal{M}$, with the topology it inherits from $\mathcal{M}$.

It is now easy to define the Martin boundary, or rather the Martin completion of $\mathcal{E}$. Under our hypotheses the map $x \to U_x$ is a continuous injection of $\mathcal{E}$ into $\mathcal{M}$, and the images of distinct points lie on distinct rays. We shall suppose this map to be in fact a homeomorphism of $\mathcal{E}$ on a subspace of $\mathcal{M}$. (This hypothesis serves only to simplify the language.) The map $x \to U_x$ composed with the natural map of $\mathcal{M}$ on $\mathcal{M}^*$ gives a homeomorphism of $\mathcal{E}$ on part of $\mathcal{M}^*$, and from now on we shall identify $\mathcal{E}$ with its image. The

Martin completion $\mathcal{E}^*$ is then the closure of $\mathcal{E}$ in $\mathcal{M}^*$, while the Martin entrance boundary is $\mathcal{E}^* - \mathcal{E}$.

In the context of the entrance boundary it is desirable to enlarge the current notion of Markoff process, as in [2], by permitting spaces of infinite measure and providing an initial as well as a terminal time. To be precise, we shall understand by Markoff process a triple (X, S, T) with the following properties.

There is a measure space $(\Omega, \mathcal{P})$, perhaps of infinite measure. S and T are measurable real functions on Ω satisfying the inequality $S < T$. The function $X = X(t, \omega)$ is defined for ω in Ω and t in the open interval $(S(\omega), T(\omega))$, and its values lie in $\mathcal{E}$. For s real and E a compact subset of $\mathcal{E}$, the subset $\Lambda_{s,E}$ of Ω defined by the conditions

$$S(\omega) < s < T(\omega), \quad X(s, \omega) \in E,$$

is a measurable set of finite measure. Finally, denote by T' the restriction of T to $\Lambda_{s,E}$ and by X' the restriction of X to the set of pairs (t, ω) for which $\omega \in \Lambda_{s,E}$ and $s \leqslant t < T(\omega)$; then (X', T') must be a Markoff process in the usual sense, having $\Lambda_{t,E}$ for basic measure space, T' for terminal time, and $P_t(x, dy)$ for transition function.

In view of the last condition, the only significant extension is the introduction of an initial time, which suggests studying the initial behavior of a Markoff process.

A Markoff process (X, S, T) determines a measure ζ on $\mathcal{E}$ by the formula

$$\zeta(E) = \int_\Omega P(d\omega) \int_{S(\omega)}^{T(\omega)} \chi_E(X(t, \omega))\, dt, \tag{4}$$

where χ_E is the characteristic function of the set E. In order to temper the initial behavior of the process we shall consider only processes for which ζ is finite on finite sets. Under this restriction ζ is excessive.

Given a Markoff process (X, S, T) we may regard $X(t, \omega)$ as a point of $\mathcal{E}^*$, using the imbedding of $\mathcal{E}$ in $\mathcal{E}^*$ described above. We shall prove that, for almost all ω, the point $X(t, \omega)$ has a limit in $\mathcal{E}^*$ as t decreases to $S(\omega)$. This statement, in itself unexciting, contains in germ most of what is known about the relations between the Martin boundary and Markoff processes.

Let us first consider a simple situation in which no difficulties arise. Assume that $\mathcal{P}(\Omega)$ is finite and that $U_x h$ is identically 1 for some strictly positive continuous function h on $\mathcal{E}$. Take Δ to be the set of all excessive measures α that satisfy

$$\int_{\mathcal{E}} h(x)\, \alpha(dx) \leqslant 1. \tag{5}$$

It is in this compact set, rather than in $\mathcal{E}^*$ itself, that the discussion of convergence will take place. We identify $\mathcal{E}$ with a subset of Δ by the map $x \to U_x$; extending this notation, we shall at times denote an element α of Δ by U_α.

If f is a positive continuous function on $\mathcal{E}$ with compact support, the function $F(x) = U_x f$ is bounded and dominates $P_t F$ for each t. Therefore the random variables

$$\varphi(t, f, \omega) = \begin{cases} \sup F, & t \leqslant S(\omega), \\ F(X(t, \omega)), & S(\omega) < t < T(\omega), \\ 0, & t \geqslant T(\omega), \end{cases}$$

form a bounded supermartingale. Since Ω has finite measure, martingale theory ensures the existence of the limit

$$\eta_\omega f = \lim_{t \searrow S(\omega)} \varphi(t, f, \omega), \tag{6}$$

for almost all ω. Using (6) for a properly chosen countable family of functions f, we find that for almost all ω the measure $U_{X(t,\omega)}$ converges to an element η_ω of Δ.

In order to transfer the convergence from Δ to $\mathcal{E}^*$ we must prove that η_ω is almost certainly not the zero measure; we shall in fact establish

$$\int_{\mathcal{E}} h(x)\, \eta_\omega(dx) = 1 \quad \text{for almost all } \omega. \tag{7}$$

To simplify the proof we shall suppose S to be identically $-\infty$. Denote by μ the distribution measure of the random point η_ω of Δ, by Ω_t the set on which T exceeds t, by μ_t the distribution measure of $X(t, \omega)$ regarded as a random point of Δ defined over Ω_t. By what has been proved, μ_t tends weakly to μ as $t \to -\infty$, so that

$$\int_\Delta \mu_t(d\alpha) \int_{\mathcal{E}} U_\alpha(dx)\, f(x) \to \int_\Delta \mu(d\alpha) \int_{\mathcal{E}} U_\alpha(dx)\, f(x) \tag{8}$$

for every function f which is continuous on $\mathcal{E}$ and has compact support, for then $U_\alpha f$ is continuous in α. The convergence

$$\int_\Delta \mu_t(d\alpha)\, U_\alpha \nearrow \zeta \qquad (t \to -\infty), \tag{9}$$

is also immediate; one has only to observe that a formula like (4) relates the measure $\int U_\alpha \mu_t(d\alpha)$ to the process obtained from (X, S, T) by replacing S by t and restricting the underlying measure space to Ω_t. From (8) and (9) follows the representation

$$\zeta = \int_\Delta U_\alpha \mu(d\alpha), \tag{10}$$

to which we shall return; at the moment we note only that (10) implies

$$\int_{\mathcal{E}} h\, d\zeta = \int_\Delta U_\alpha h \mu(d\alpha). \tag{11}$$

Since μ_t is concentrated on $\mathcal{E}$ and $U_x h = 1$ for every x in $\mathcal{E}$, relation (9) implies

$$\mu_t(\Delta) = \int_\Delta U_\alpha h \mu_t(d\alpha) \to \int_{\mathcal{E}} h\, d\zeta,$$

so that $\int h\,d\zeta$ is just $\mu(\Delta)$. This equality, relation (11) and the fact that $U_\alpha h$ cannot exceed 1 together imply that $U_\alpha h = 1$ for all α in Δ, excepting a set of μ-measure zero. Thus (7) holds, and the proof is complete for the simple situation. It should be noted that the proof does not involve the hypotheses we have made about the transition function; it is valid when $\mathcal{E}$ can be properly imbedded in the Martin completion and the process has sufficiently regular sample paths. This remark will be needed for the application we are about to make.

Let now (X, S, T) be an arbitrary Markoff process, subject to the restriction that the measure ζ is finite on compact sets. We shall employ a transformation studied by Dynkin [1] in another context to reduce the situation to the simple one. Choose a strictly positive continuous function g so that $\int g\,d\zeta$ is finite and $G(x) = U_x g$ is continuous in x; the existence of g is ensured by our hypotheses on $P_t(x, dy)$, and G is clearly strictly positive. Define a new transition function by the formula

$$P'_t(x, dy) = \frac{1}{G(x)} P_t(x, dy)\, G(y).$$

The corresponding kernel is then

$$U'_x(dy) = \frac{1}{G(x)} U_x(dy)\, G(y).$$

According to these formulas the Martin completion of $\mathcal{E}$ determined by $P'_t(x, dy)$ may be identified with the one determined by $P_t(x, dy)$.

Construct also a new process (X', S', T') over a measure space $(\Omega', \mathcal{P}')$. The set Ω' is the part of $\Omega \times R$ comprising the points $\omega' = (\omega, \varrho)$ for which $S(\omega) < \varrho < T(\omega)$, and $\mathcal{P}'$ is the measure

$$\mathcal{P}'(d\omega') = \mathcal{P}'(d\omega, d\varrho) = g(X(\varrho, \omega))\, \mathcal{P}(d\omega)\, d\varrho,$$

where $d\varrho$ stands for Lebesgue measure. By (4), the total mass of Ω' is

$$\mathcal{P}'(\Omega') = \int_{\mathcal{E}} g\,d\zeta,$$

which is finite. Finally, set

$$S(\omega') = S(\omega), \quad T(\omega') = \varrho, \quad X'(\omega', t) = X(\omega, t),$$

for $\omega' = (\omega, \varrho)$. Evidently the sample paths of (X', S', T') begin in just the same manner as those of the original process (X, S, T).

One easily verifies that (X', S', T') is a Markoff process corresponding to the transition function $P'_t(x, dy)$ and that the measure related to it by (4) is $G(x)\,\zeta(dx)$. Thus we have only to take h' to be g/G to show that we have reached the simple situation.

Consequently, for almost all ω', the point $X'(t, \omega')$ has a limit in $\mathcal{E}^*$ as t decreases, and on reviewing the definitions we obtain the corresponding statement for (X, S, T).

The convergence being established, one can use standard arguments to explore the relations between Markoff processes and the Martin completion.

It is not hard to prove, for example, that in the representation (10) of ζ almost every measure U_α is an extreme point of Δ, and that ζ has essentially just one such representation by extreme excessive measures. In this way one obtains the analogue of the Poisson–Riesz–Martin representation for excessive measures that arise from a Markoff process; unfortunately, one must enlarge still further the notion of Markoff process to obtain the result for all excessive measures.

REFERENCES

[1]. Dynkin, E. B., Transformations of Markov processes connected with additive functionals. *Proc. Fourth Berkeley Symp. on Stat. and Prob.* 1960, 117–142.

[2]. Hunt, G. A., Markoff chains and Martin boundaries. *Illinois J. Math.*, 4 (1960), 313–340.

[3]. Ray, D., Resolvents, transition functions, and strongly Markovian processes. *Ann. Math.*, 70 (1959), 43–71.

THE BROWNIAN MOTION AND TENSOR FIELDS ON RIEMANNIAN MANIFOLD

By KIYOSI ITÔ

1. Introduction

The Brownian motion $\xi(t)$ on an r-dimensional Riemannian manifold M^r is defined as a diffusion process with the generator

$$g = \tfrac{1}{2} g^{ij} \nabla_i \nabla_j, \tag{1.1}$$

where ∇_i denotes covariant differentiation. The paths of the Brownian motion can be constructed by means of the stochastic differential equation [1,2,3]. We shall denote with P_a the probability law of the Brownian motion starting at a point $a \in M^r$ and the corresponding expectation with E_a.

Let f be a scalar field on M^r. It is clear by the definition that

$$u(t,a) = E_a[f(\xi(t))] \tag{1.2}$$

satisfies a heat equation.

$$\frac{\partial}{\partial t} u = \tfrac{1}{2} g^{ij} \nabla_i \nabla_j u \tag{1.3 a}$$

with the initial condition

$$u(0+,a) = f(a). \tag{1.3}$$

Our problem is to establish a similar fact in case f is a vector field f_k or more generally a tensor field $f_{k_1 k_2 \dots k_p}$. We cannot replace f in (1.2) with $f_{k_1 k_2 \dots k_p}$, because $f_{k_1 k_2 \dots k_p}(\xi(t))$ is a tensor attached to the point $\xi(t)$ which varies with t, while $u(t,a)$ should be attached to $a = \xi(0)$. Therefore we shall shift $f_{k_1 \dots k_p}(\xi(t))$ back to $a = \xi(0)$ along the path ξ by parallel displacement. Denoting with $T_{\xi,t}$ the parallel displacement along ξ from $a = \xi(0)$ to $\xi(t)$, we can prove that

$$u_{k_1 \dots k_p}(t,a) = E_a[T_{\xi,t}^{-1} f_{k_1 \dots k_p}(\xi(t))] \tag{1.4}$$

satisfies a differential equation of the same form as (1.3):

$$\frac{\partial}{\partial t} u_{k_1 \dots k_p} = \tfrac{1}{2} g^{ij} \nabla_i \nabla_j u_{k_1 \dots k_p}. \tag{1.5}$$

In order to define the parallel displacement $T_{\xi,t}$ used above, we cannot use Levi-Civita's definition in its usual form because of the non-differentiability of the Brownian paths. We shall define $T_{\xi,t}$ as follows. Let

$$\Delta : 0 = t_0 < t_1 < \dots < t_n = t \tag{1.6}$$

be a division of the time interval $[0,t]$, make a polygonal curve $\xi_\Delta(t)$ which consists of the geodesic curves connecting $\xi(t_{i-1})$ with $\xi(t_i)$, $i=1,2,\ldots,n$, and denote with $T_{\xi,t}$ the Levi-Civita parallel displacement along ξ_Δ. Then we can prove that $T_{\xi_\Delta,t}$ tends to a limit in probability as $|\Delta|=\max_i(t_i-t_{i-1})$ tends to 0. This limit is defined to be $T_{\xi,t}$.

Noticing that the operator $\Delta=d\delta+\delta d$ introduced in the Hodge theory of harmonic tensor fields by Kodaira and also by Bidal and de Rham is expressed in the form [5]

$$\Delta=-g^{ij}\nabla_i\nabla_j+\text{linear transformation} \tag{1.7}$$

and modifying the transformation $T_{\xi,t}$ slightly in the above discussion, we can also get a solution of

$$\frac{\partial}{\partial t}u_{k_1\ldots k_p}=-\tfrac{1}{2}\Delta u_{k_1\ldots k_p} \tag{1.8 a}$$

with the initial condition

$$u_{k_1\ldots k_p}(0+,a)=f_{k_1\ldots k_p}(a) \tag{1.8 b}$$

As was proved by Milgram and Rosenbloom [4], $u_{k_1\ldots k_p}(+\infty,a)$ is the harmonic tensor field with the same periods as $f_{k_1\ldots k_p}(a)$ in case M^r is compact. We shall discuss this in a separate paper.

2. The paths of the Brownian motion on a Riemannian manifold

The Brownian motion on a Riemannian manifold M^r is a diffusion (*a* strict Markov process with continuous paths) with the generator

$$\begin{aligned} g&=\tfrac{1}{2}g^{ij}\nabla_i\nabla_j\\ &=\tfrac{1}{2}g^{ij}\frac{\partial^2}{\partial x^i\partial x^j}-\tfrac{1}{2}g^{ij}\Gamma^k_{ij}\frac{\partial}{\partial x^k}. \end{aligned} \tag{2.1}$$

Introducing σ^i_k and m^k by

$$\sum_k\sigma^i_k\sigma^j_k=g^{ij}, \tag{2.2 a}$$

$$m^k=-\tfrac{1}{2}g^{ij}\Gamma^k_{ij} \tag{2.2 b}$$

and solving a stochactic differential equation

$$d\xi^i(t)=\sigma^i_k(\xi(t))\,d\beta^k(t)+m^i(\xi(t))\,dt, \tag{2.3 a}$$

$$\xi^i(0)=a^i \tag{2.3 b}$$

(β is the n-dimensional Wiener process),

we can construct the paths of the Brownian notion starting at $a\in M^r$. See [1, 2 and 3] for the details.

3. The parallel displacement along the Brownian paths

Since almost all paths are non-differentiable, we shall define the parallel displacement along the Brownian paths as follows.

Let $(\xi^k(t),\ k=1,2,\dots,r)$ be the Brownian notion on M^r starting at $a=(a^k)$ and $\alpha_{k_1 k_2 \dots k_p}$ be a tensor attached to a. Let

$$\Delta: 0=t_0<t_1<t_2<\dots<t_n=t \tag{3.1}$$

be a division of the time interval $[0,t]$, make a polygonal curve $\xi_\Delta(s), 0\leqslant s\leqslant t$, which consists of the geodesic curves connecting $\xi(t_{i-1})$ with $\xi(t_i)$, $i=1, 2,\dots,n$, and denote with $\alpha^\Delta_{k_1\dots k_p}(t)$ be the tensor obtained from $\alpha_{k_1\dots k_p}$ by the Levi-Civita parallel displacement along ξ_Δ. Then we have

THEOREM 1. *As* $|\Delta|=\max_i(t_i-t_{i-1})$ *tends to* 0, $\alpha^\Delta_{i_1\dots i_p}(t)$ *tends in probability for each fixed* t *to the solution* $\alpha_{i_1\dots i_p}(t)$ *of the stochastic differential equation*

$$\begin{aligned}
d\alpha_{k_1\dots k_p}(t)
&=\sum_\nu \Gamma^k_{ik_\nu}(\xi(t))\,\alpha_{k_1\dots k_{\nu-1}kk_{\nu+1}\dots k_p}(t)\,d\xi^i(t)\\
&\quad+\tfrac12\sum_\nu g^{mi}(\xi(t))\left[\frac{\partial\Gamma^k_{ik_\nu}}{\partial x^m}(\xi(t)+\Gamma^l_{ik_\nu}(\xi(t))\,\Gamma^k_{ml}(\xi(t))\right]\\
&\qquad\times\alpha_{k_1\dots k_{\nu-1}kk_{\nu+1}\dots k_p}(\xi(t))\,dt\\
&\quad+\tfrac12\sum_{\mu\neq\nu} g^{mi}(\xi(t))\,\Gamma^k_{mk_\mu}(\xi(t))\,\Gamma^l_{ik_\nu}(\xi(t))\\
&\qquad\times\alpha_{k_1\dots k_{\mu-1}kk_{\mu+1}\dots k_{\nu-1}lk_{\nu+1}\dots k_p}(\xi(t))\,dt,
\end{aligned}\tag{3.2 a}$$

with the initial condition

$$\alpha_{k_1\dots k_p}(0)=\alpha_{k_1\dots k_p}. \tag{3.2 b}$$

See [1,2] for the meaning of the this equation.

Definition. $\alpha_{k_1\dots k_p}(t)$ is called the parallel displacement of $\alpha_{k_1\dots k_p}$ along ξ.

We shall sketch the proof of this theorem.

Noticing the symbolic estimation

$$\Delta\xi^i(t)\,\Delta\xi^j(t)\sim g^{ij}(\xi(t))\,\Delta t \tag{3.3}$$

(see [2] for the exact meaning), we can ignore $o((\Delta\xi^i(t))^2)$ as well as $o(\Delta t)$. Therefore the geodesic curve C connecting $a=\xi(t)$ with $b=\xi(t+\Delta t)$ is approximately

$$x^i(t)=a^i+(b^i-a^i)t+\frac{t(1-t)}{2}\Gamma^i_{jk}(a)\,(b^j-a^j)\,(b^k-a^k). \tag{3.4}$$

Therefore the tensor $\beta_{k_1\dots k_p}$ at b obtained from $\alpha_{k_1\dots k_p}$ at a by the Levi-Civita parallel displacement along C is approximately given by

$$\beta_{k_1\dots k_p} \sim \alpha_{k_1\dots k_p} + \sum_\nu \Gamma^k_{ik_\nu}(a)\alpha_{k_1\dots k_{\nu-1}kk_{\nu+1}\dots k_p}(b^i - a^i)$$
$$+ \frac{1}{2}\frac{\partial \Gamma^k_{ik_\nu}}{\partial x^m}(a) + \Gamma^l_{ik_\nu}(a)\Gamma^k_{ml}(a)$$
$$\times \alpha_{k_1\dots k_{\nu-1}kk_{\nu+1}\dots k_p}(a)(b^m - a^m)(b^i - a^i)$$
$$+ \tfrac{1}{2}\sum_{\mu\neq\nu} \Gamma^l_{ik_\nu}(a)\Gamma^l_{mk_\mu}(a)\alpha_{i_1\dots i_{\mu-1}ki_{\mu+1}\dots i_{\nu-1}li_{\nu+1}\dots i_p}(a)$$
$$\times (b^m - a^m)(b^i - a^i). \tag{3.5}$$

using (3.3) we shall get the stochastic differential equation (3.2 a) for $\alpha_{i_1\dots i_p}(t)$.

4. Heat equation

Using the notations as in section 3, consider the mapping

$$T_{\xi,t} : \alpha_{k_1\dots k_p} \to \alpha_{k_1\dots k_p}(t) \tag{4.1}$$

and the tensor

$$u_{k_1\dots k_p}(t, a) = E_a(T^{-1}_{\xi,t} f_{k_1\dots k_p}(\xi(t))) \tag{4.2}$$

Then we have

THEOREM 2. $u_{k_1\dots k_p}(t, a)$ satisfies the heat equation

$$\frac{\partial}{\partial t} u_{k_1\dots k_p} = \tfrac{1}{2} g^{ij} \nabla_i \nabla_j u_{k_1\dots k_p}. \tag{4.3}$$

The Markov property of the Brownian motion implies that

$$\frac{\partial}{\partial t} u_{k_1\dots k_p} = A u_{k_1\dots k_p}, \tag{4.4}$$

where
$$A f_{k_1\dots k_p} = \lim_{t\to 0} \frac{f_{k_1\dots k_p}(t, a) - f_{k_1\dots k_p}(a)}{t}. \tag{4.5}$$

To identify A with $\frac{1}{2}g^{ij}\nabla_i\nabla_j$, it is enough to compute $E(T^{-1}_{\xi,t} f_{i_1\dots i_p}(\xi(t))$ neglecting $o((d\xi^i)^2)$ as well as $o(dt)$. Though the computation is complicated, the technique is essentially the same as in [1].

REFERENCES

[1]. ITÔ, K., Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.*, 1 (1950), 35–47.

[2]. —— On a formula concerning stochastic differentials. *Nagoya Math. J.*, 3 (1951), 55–65.

[3]. —— Stochastic differential equations in a differentiable manifold. *Mem. Coll. Sci. Univ. Kyoto, Ser. A. Math.*, 28, (1953), 81–85.

[4]. MILGRAM, A. N. & ROSENBLOOM, P. C., Harmonic forms and heat conduction. I. Closed Riemannian manifolds. *Proc. Nat. Acad. Sci. U.S.A.*, 37 (1951), 180–184.

[5]. DE RHAM, G., Variétés différentiables. *Actualités Sci. Ind.*, 1222 (1955), Paris.

ВЕРОЯТНОСТНЫЕ ИДЕИ В ЭРГОДИЧЕСКОЙ ТЕОРИИ

Я. Г. СИНАЙ

В настоящее время происходит весьма интенсивное развитие эргодической теории. В предлагаемом докладе будет затронута часть последних достижений в этой области, связанная с новыми идеями и методами, пришедшими из теории вероятностей. Первой работой, где эти новые идеи и методы появились, была работа А. Н. Колмогорова [26]. Понятие энтропии динамической системы и понятие K-системы, введенные в [26], заняли в настоящее время прочное место в метрической теории динамических систем. Сейчас уже ясно, что происходящий на наших глазах прогресс обязан своим существованием не столько появлению численной характеристики динамической системы — энтропии, сколько выделению и исследованию роли так называемых возрастающих σ-алгебр или, что то же, возрастающих разбиений (см. § 1). Последние представляют собой, по существу, результат переработки для сохраняющих меру преобразований идей теории вероятностей, применявшихся при изучении вопросов регулярности и сингулярности стационарных случайных процессов. Старые понятия и представления приобрели при этом новый смысл, оказавшись чрезвычайно полезными в целом ряде старых проблем эргодической теории. Мы имеем в настоящий момент довольно стройную теорию, далекую от своего завершения, но уже установившую прочные связи с функциональным анализом, вариационным исчислением, классической механикой, дифференциальной геометрией и т. д.

В обстоятельном обзоре В. А. Рохлина [38] изложены результаты, полученные после появления работы [26] до 1960 г. С тех пор появился ряд новых результатов и применений теории. Некоторые из них, и притом далеко не все, излагаются далее.

§ 1. Возрастающие подалгебры и K-системы

Метрическая теория динамических систем предполагает заданным пространство с мерой, т. е. множество M, некоторую σ-алгебру $\mathfrak{S}$ его подмножеств и меру μ, определенную на элементах из $\mathfrak{S}$. Обычно ограничиваются случаем, когда M есть пространство Лебега с непрерывной мерой, т. е. изоморфно mod 0 единичному отрезку с обычной мерой Лебега. Теория пространств Лебега изложена в работе В. А. Рохлина [34]. Чрезвычайно важным является факт эквивалентности понятия измеримого разбиения, центрального в теории пространств Лебега, и понятия подалгебры алгебры $\mathfrak{S}$: при помощи измеримого разбиения ζ строится подалгебра $\mathfrak{M}(\zeta)$ σ-алгебры $\mathfrak{S}$, образованная множествами, составленными mod 0 из элементов разбиения ζ, и, обратно, для всякой подалгебры $\mathfrak{M}$ σ-алгебры $\mathfrak{S}$ существует единственное mod 0 разбиение ζ, для которого $\mathfrak{M}=\mathfrak{M}(\zeta)$.

В этом параграфе мы будем иметь дело в начале с отдельными автоморфизмами пространства M, т. е. с взаимооднозначными преобразованиями пространства M на себя, сохраняющими меру μ. Многочисленные примеры таких автоморфизмов доставляет теория вероятностей: для каждого стационарного (в узком смысле) процесса $\xi=\{\xi_n\}$, $-\infty<n<\infty$ сдвиг $T: T\xi=\xi^1=\{\xi_n^1\}$, $\xi_n^1=\xi_{n-1}$ является автоморфизмом пространства реализаций процесса ξ.

Подалгебру $\mathfrak{M}$ алгебры $\mathfrak{S}$ будем называть возрастающей относительно автоморфизма T, если

$$T\mathfrak{M}\supset\mathfrak{M} \quad \text{(строгое включение).}$$

Из одного только факта существования возрастающей подалгебры $\mathfrak{M}$ можно сделать ряд заключений о спектре унитарного оператора U, сопряженного с автоморфизмом T, и о частичном перемешивании множеств из $\mathfrak{S}$ (см. [38], [47]).

В том частном случае, когда автоморфизм T представлен как сдвиг реализации некоторого стационарного процесса $\xi=\{\xi_n\}$, естественно рассмотреть в качестве подалгебры $\mathfrak{M}$ подалгебру, порожденную случайными величинами ξ_i при $i\leqslant 0$. Очевидно, что для нее всегда $T\mathfrak{M}\supseteq\mathfrak{M}$. Условие $T\mathfrak{M}\supset\mathfrak{M}$ выделяет класс хорошо известных в теории вероятностей несингулярных стационарных процессов, способных передавать во времени информацию. Наоборот, сингулярные процессы определяются равенством $T\mathfrak{M}=\mathfrak{M}$. Существенно, однако, что и в случае сингулярных процессов возможно, что у соответствующего автоморфизма существуют возрастающие подалгебры. Интересным примером может служить сдвиг гауссовского стационарного процесса, у которого спектральная плотность (по лебеговской мере) обращается в нуль на некотором отрезке (см. [41]). Важно отметить что подобные примеры невозможны, если ξ процесс с конечным числом состояний: из равенства $T\mathfrak{M}=\mathfrak{M}$ вытекает в этом случае, что у автоморфизма T возрастающих подалгебр вообще существовать не может.

Первый шаг в изучении возрастающих подалгебр состоит в выделении сингулярной части автоморфизма T. Предыдущее замечание подсказывает путь, следуя которому это можно сделать достаточно инвариантным образом.

Пусть $\mathfrak{A}$—конечная подалгебра. Полагаем

$$\mathfrak{A}^-=\bigvee_{k\leqslant 0}T^k\mathfrak{A}, \quad \underline{\mathfrak{M}}=\bigvee_{\mathfrak{A}}\bigwedge_k T^k\mathfrak{A}^-$$

($\vee$ и $\wedge$ — объединение и пересечение подалгебр). Подалгебра $\underline{\mathfrak{M}}$ — инвариантна, т. е. $T\underline{\mathfrak{M}}=\underline{\mathfrak{M}}$. Кроме того, для любой подалгебры $\mathfrak{M}_1\subseteq\underline{\mathfrak{M}}$ из включения $T\mathfrak{M}_1\supseteq\mathfrak{M}_1$ вытекает равенство $T\mathfrak{M}_1=\mathfrak{M}_1$.

Если автоморфизм представлен как сдвиг реализации стационарного случайного процесса ξ с конючным числом состояний, то подалгебру $\underline{\mathfrak{M}}$ можно получить пересечением

$$\underline{\mathfrak{M}}=\bigwedge_k T^k\mathfrak{M},$$

где $\mathfrak{M}$ — подалгебра, порожденная случайными величинами ξ_i при $i\leqslant 0$. Подалгебра $\underline{\mathfrak{M}}$ была введена М. С. Пинскером [32].

В дальнейшем мы исключим из рассмотрения автоморфизмы T, для которых $\underline{\mathfrak{M}}=\mathfrak{S}$, или, что то же, класс автоморфизмов, у которых отсутствуют возрастающие подалгебры (такие автоморфизмы можно еще охарактеризовать равенством нулю их энтропии (см. ниже)).

Возрастающая подалгебра $\mathfrak{M}$ называется простой, если

$$\bigwedge_k T^k\mathfrak{M}\subseteq\underline{\mathfrak{M}}.$$

Методами работы [41] можно показать, что всякая возрастающая подалгебра $\mathfrak{M}$ содержит такую простую подалгебру $\mathfrak{M}_1$, что $\bigvee_k T^k\mathfrak{M}=\bigvee_k T^k\mathfrak{M}_1$. Это значит, что из подалгебры $\mathfrak{M}$ можно выбросить «лишние множества», которые более естественно получать в процессе сдвига меньшей подалгебры $\mathfrak{M}_1$, чем иметь в самой алгебре $\mathfrak{M}$. Благодаря этому оказывается, что при изучении произвольных возрастающих подалгебр можно ограничиться простыми подалгебрами.

Среди простых подалгебр $\mathfrak{M}$ всегда существуют (см. [41]) экстремальные подалгебры. Так мы называем возрастающие подалгебры $\mathfrak{M}$, для которых

$$\bigwedge_k T^k\mathfrak{M}=\underline{\mathfrak{M}},\qquad \bigvee_k T^k\mathfrak{M}=\mathfrak{S}.$$

Изучение экстремальных подалгебр связано с тонкими вопросами теории пересечений σ-алгебр. Не входя в подробности, приведем основные факты.

Экстремальная подалгебра $\mathfrak{M}_1$ называется совершенной, если для любой экстремальной подалгебры $\mathfrak{M}\subset\mathfrak{M}_1$

$$\bigwedge_k (T^k\mathfrak{M}_1\vee\mathfrak{M})=\mathfrak{M}.$$

Важной характеристикой экстремальной подалгебры $\mathfrak{M}$ является условная энтропия $H(T\mathfrak{M}\,|\,\mathfrak{M})$. Известны примеры автоморфизмов, у которых имеются екстремальные подалгебры с разным значением этой величины [26]. Заметим в этой связи, что для такого классического примера автоморфизма, как сдвиг последовательности $\xi=\{\xi_n\}$ независимых равномерно распределеных на отрезке [0, 1] случайных величин для любого $=h$, $0<h\leqslant\infty$ для которого число 2^h является трансцендентным, можно найти экстремальную подалгебру $\mathfrak{M}$, для которой

$$H(T\mathfrak{M}\,|\,\mathfrak{M})=h.$$

Совершенные подалгебры характеризуются тем, что для них величина $H(T\mathfrak{M}\,|\,\mathfrak{M})$ принимает самое большое возможное значение. Обозначая его через $h(T)$, получим то, что называется энтропией автоморфизма T. В случае $h(T)<\infty$ равенство $H(T\mathfrak{M}\,|\,\mathfrak{M})=h(T)$ необходимо и достаточно для того, чтобы экстремальная подалгебра была совершенной. В случае $h(T)=\infty$ нетрудно привести примеры, когда это уже не так. Совершенные подалгебры следует считать наиболее естественным аналогом уже упоминавшейся, принятой в теории вероятностей подалгебры событий, зависящих от течения стационарного процесса до момента времени 0.

Известная проблема эргодической теории — проблема существования счетного образующего разбиения у автоморфизма T (см. [38]).

Последняя может быть сформулирована также как проблема представления автоморфизма T в виде сдвига реализации некоторого стационарного процесса с не более чем счетным числом состояний. Факт существования экстремальной подалгебры показывает, что нечто, присущее такому представлению, сопутствует автоморфизму T всегда — это аналог подалгебры событий, зависящих от поведений процесса до момента времени 0.

Важнейший класс автоморфизмов определяется требованием $\underline{\mathfrak{M}}=\mathfrak{N}$, где $\mathfrak{N}$ — тривиальная σ-алгебра из множеств меры нуль или единица. В этом случае автоморфизм T называется K-автоморфизмом. Из предыдущего ясно, что экстремальная подалгебра $\mathfrak{M}$ для K-автоморфизма удовлетворяет соотношениям

$$\bigvee_k T^k\mathfrak{M}=\mathfrak{S}, \quad \bigwedge_k T^k\mathfrak{M}=\mathfrak{N}.$$

Требование существования возрастающего разбиения $\mathfrak{M}$ с такими двумя свойствами можно взять за еще одно определение K-автоморфизма [26]. Когда автоморфизм T есть сдвиг реализации стационарного процесса ξ, равенство

$$\bigwedge_k T^k\mathfrak{M}=\mathfrak{N}$$

для подалгебры $\mathfrak{M}$, порожденной случайными величинами ξ_i при $i\leqslant 0$, превращается в определение регулярного процесса или, как еще принято говорить, процесса, удовлетворяющего закону «нуля или единицы». Таким образом, K-автоморфизмы — это аналоги регулярных стационарных процессов.

K-автоморфизмы эргодичны. Унитарный оператор, сопряженный с K-автоморфизмом, имеет в ортогональном дополнении к подпространству постоянных счетнократный лебеговский спектр. Автоморфизм, обратный к K-автоморфизму, и всякий фактор-автоморфизм K-автоморфизма снова являются K-автоморфизмами. Интересны свойства перемешивания, которыми обладают K-автоморфизмы. Пусть A и B — два произвольных множества из γ. В случае простого перемешивания

$$\lim_{n\to\infty}\mu(A\cap T^nB)=\mu(A)\,\mu(B).$$

Если есть кратное перемешивание степени k, то

$$\lim_{\substack{n_1\to\infty\\ n_2\to\infty\\ \cdots\cdots\\ n_k\to\infty}}\mu(A\cap T^{n_1}B\cap T^{n_1+n_2}B\cap\ldots\cap T^{n_1+n_2n+\ldots+n_k}B)=\mu(A)\,[\mu(B)]^{k-1}.$$

Если же T является K-автоморфизмом, то оба предыдущих соотношения могут быть значительно усилены. Обозначим через $\mathfrak{S}^k$ σ-алгебру, порожденную множествами T^nB, $n\geqslant k$. Тогда оказывается, что

$$\lim_{k\to\infty}\,[\sup_{C\in\mathfrak{S}^k}|\mu(A\cap C)-\mu(A)\,\mu(C)|]=0. \tag{1}$$

Такое равномерное перемешивание означает, конечно, гораздо более

сильное условие, чем обычное перемешивание. Оно может даже служить определением K-автоморфизма: всякий автоморфизм T, для которого при любых множествах A и B выполнено (1), является K-автоморфизмом. Это легко вытекает из результатов работ [47], [41].

В теории стационарных процессов известен ряд усилений понятия регулярности: понятие «сильного перемешивания» Розенблатта [64], понятие информационной регулярности и т. д. (см. [15]). Неясно, однако, существует ли для них достаточно инвариантные определения, которое сделали бы их полезными для эргодической теории.

В приложениях эргодической теории к классическим динамическим системам наиболее важен случай потоков, т. е. непрерывных однопараметрических групп автоморфизмов. Теории, аналогичной изложенной выше для одного автоморфизма, здесь еще нет. Приведем те факты, которые пока известны. Пусть $\{S_t\}$ — поток. Для каждого отдельного автоморфизма S_τ можно построить подалгебру $\underline{\mathfrak{M}}$. Нетрудно показать, что эта подалгебра $\underline{\mathfrak{M}}$ будет одна и та же для всех автоморфизмов S_τ. Основная трудность состоит в том, что неизвестно, вытекает ли теперь из неравенства $\underline{\mathfrak{M}} \neq \mathfrak{S}$ существование возрастающей подалгебры для потока, т. е. такой подалгебры $\mathfrak{M}$, для которой

$$S_t\mathfrak{M} \supset \mathfrak{M}$$

при всех $t>0$. Ответ на этот вопрос неизвестен даже в случае $\mathfrak{M}=\mathfrak{N}$. Частичные результаты в этом направлении были получены Б. М. Гуревичем (не опубликовано). Следует подчеркнуть, что трудности, которые здесь имеются, похожи на трудности, встречающиеся в теории информации в вопросах определения количества информации, создаваемого в одном процессе о другом, за единицу времени (см. [33].

Поток $\{S_t\}$ называется теперь K-потоком (см. [26]), если существует подалгебра $\mathfrak{M}$, удовлетворяющая условиям:

$$1)\ S_t\mathfrak{M} \supseteq \mathfrak{M} \text{ при } i>0; \quad 2)\ \bigvee_t S_t\mathfrak{M} = \mathfrak{S} \text{ и } \quad 3)\ \bigwedge_t S_t\mathfrak{M} = \mathfrak{N}.$$

K-потоки эргодичны. Соответствующая K-потоку непрерывная группа унитарных операторов имеет в ортогональном дополнении к подпространству постоянных счетнократный лебеговский спектр (см. [47]). Всякий автоморфизм S_τ, входящий в K-поток, является K-автоморфизмом. Поэтому для K-потока можно сформулировать свойства перемешивания, аналогичные (1). Неизвестно, однако, могут ли они теперь заменить собой определение K-потока.

§ 2. Слабый изоморфизм

Появление энтропии вновь сделало острой проблему изоморфизма динамических систем. Около тридцати лет назад von Neumann решил ее для динамических систем с дискретным спектром. С тех пор более двадцати пяти лет продолжались попытки распространить результат von Neumann'а на случай систем с непрерывным спектром. Предлагались различные инварианты, строились интересные примеры, но проблема так и осталась открытой. Наконец, существенный сдвиг был сделан А. Н. Колмогоровым [26] введением энтропии.

Ставшее сейчас уже традиционным введение энтропии осуществляют, исходя из конечных или счетных разбиений пространства M [37], [42]. Пусть η-разбиение пространства M на не более чем счетное число множеств положительной меры B_1, B_2, ... Полагаем

$$H(\eta) = -\sum_i \mu(B_i) \log \mu(B_i).$$

Класс разбиений η, для которых $H(\eta) < \infty$, обозначим через Z. Возьмем разбиение η^n пространства M на множества $TB_i, \cap T^2 B_{i_2} \cap \ldots \cap T^n B_{i_n}$. Из теории информации известно, что существует предел $h(T, \eta)$

$$h(T, \eta) = \lim_{n \to \infty} \frac{1}{n} H(\eta^n)).$$

Верхняя грань $\sup\limits_{\eta \in Z} h(T, \eta)$ и есть энтропия $h(T)$ автоморфизма T.

Другой способ определения энтропии, описанный в предыдущем параграфе, приводит к тому же самому результату. Нетрудно показать, что условие $h(T) > 0$ эквивалентно условию $\underline{\mathfrak{M}} \subset \mathfrak{S}$. Таким образом, теория развитая в предыдущем параграфе, относится к автоморфизмам с положительной энтропией.

Простейшая теорема [42], позволяющая часто находить энтропию $h(T)$, состоит в том, что если $\eta \in Z$ есть образующее разбиение, то верхняя грань чисел $h(T, \eta)$ достигается на η (Напомним, что разбиение называется образующим, если σ-алгебра, порожденная множествами $T^k B_i$, $-\infty < k < \infty$, $i = 1, 2, \ldots$, совпадает с $\mathfrak{S}$). Эта теорема показывает, что хорошо известные в теории вероятностей формулы для энтропии на один шаг стационарного процесса с конечным или счетным числом состояний дают энтропию автоморфизма, порожденного этим процессом. В частности, если стационарный процесс есть цепь Маркова со стационарным распределением q_i и матрицей перехода $\|p_{ij}\|$, то сдвиг в пространстве реализаций процесса, называемый в этом случае автоморфизмом Маркова, имеет энтропию

$$h(T) = -\sum_i q_i \sum_j p_{ij} \log p_{ij}. \tag{2}$$

Если цепь Маркова есть последовательность независимых случайных величин ($p_{ij} = q_j$), то сдвиг называется автоморфизмом Бернулли и в этом случае (2) дает:

$$h(T) = -\sum_i q_i \log q_i.$$

Эргодические автоморфизмы Маркова с одним подклассом и автоморфизмы Бернулли являются классическими примерами K-автоморфизмов.

В обзоре [38] отмечено, что всякий автоморфизм T, у которого существует счетное образующее разбиение η с $h(T) = h(T, \eta) = H(\eta)$ является автоморфизмом Бернулли. Подобно этому, автоморфизм Маркова характеризуется тем, что для него существует счетное образующее разбиение η с $h(T) = h(T, \eta) = H(\eta \mid T^{-1}\eta)$.

Первое употребление энтропии, принадлежащее А. Н. Колмогорову, состояло в разложении совокупности всех автоморфизмов пространства

M на метрически неизоморфные друг другу множества автоморфизмов, различающиеся по значению энтропии. Поскольку такое разложение можно произвести и среди K-автоморфизмов, то тем самым стало очевидно, что спектральных инвариантов для проблемы изоморфизма недостаточно. В настоящее время для автоморфизмов с непрерывным спектром известны, вообще, только следующие инварианты метрического изоморфизма, позволяющие строить примеры неизоморфных между собой автоморфизмов: во-первых, спектр автоморфизма, во-вторых, сингулярная часть автоморфизма, т. е. фактор-автоморфизм, действующий на фактор-пространстве $M|\mathfrak{M}$, и, в-третьих, энтропия. Между этими инвариантами имеется некоторая связь [38], но мы на ней здесь не останавливаемся. Существенно только, что в классе K-автоморфизмов из этих трех инвариантов остается одна энтропия. Естественно пытаться выяснить, возможно ли дальнейшее разложение на метрически неизоморфные между собой подмножества совокупности K-автоморфизмов с данным значением энтропии. Работа Л. Д. Мешалкина [30] была первой, где удалось дать несколько примеров изоморфизма двух различных автоморфизмов Бернулли с одинаковой энтропией. Однако попытки применить метод работы [30] к сколько-нибудь широкому классу автоморфизмов не привели к цели. Более того, было построены примеры, где этот метод применим, но в этих примерах не удалось выяснить, приводит он к изоморфизму или к гомоморфизму. Последний факт оказался совершенно новым в этом круге вопросов. Он повлек за собой предположение, что между K-автоморфизмами с одинаковой энтропией возможны менее тесные взаимоотношения, чем изоморфизм.

Прежде, чем говорить о дальнейшем продвижении, остановимся на вероятностной стороне проблемы изоморфизма. Если измеримую функцию $f(x)$ на пространстве M трактовать как случайную величину, то последовательность случайных величин $\{f(T^n x)\}$ является стационарным процессом. В полную проблему изоморфизма входит описание всех процессов, которые можно получить при заданном автоморфизме, беря всевозможные функции f. Более частная постановка вопроса заключается в нахождении условий, при которых среди таких процессов можно найти процессы заданного вида, например, последовательности независимых случайных величин, цепи Маркова и т. п. Отметим, что близкие проблемы ставились в работах М. Розенблатта [65], [66].

Будем теперь называть автоморфизм T_1 пространства M_1 слабо изоморфным автоморфизму T_2 пространства M_2, если существует гомоморфизм U_1 пространства M_1 на M_2 и гомоморфизм U_2 пространства M_2 на M_1, такие, что

$$U_1 T_1 = T_2 U_1, \quad T_1 U_2 = U_2 T_2.$$

Для того чтобы T_1 и T_2 были изоморфными, нужно, чтобы существовал такой изоморфизм $U = U_1$, для которого $U_2 = U^{-1}$. Если автоморфизмы T_1 и T_2 представляют собой сдвиги реализаций двух различных стационарных процессов ξ' и ξ'', то слабый изоморфизм T_1 и T_2 означает, что процесс ξ' представим в виде сдвига некоторого функционала от ξ'' и, наоборот, ξ'' представим в виде сдвига некоторого функционала от процесса ξ'.

Нетрудно показать, что спектр автоморфизма, энтропия и равенство

$\underline{\mathfrak{M}}=\mathfrak{N}$ являются инвариантами слабого изоморфизма (см. [50]). Для автоморфизмов с дискретным спектром слабый изоморфизм эквивалентен изоморфизму. Однако в общем случае нет никаких оснований считать, что эти понятия совпадают.

Повидимому, в изучении K-автоморфизмов с точностью до слабого изоморфизма энтропия может иметь решающее значение. На это указывает результат, полученный в работе [50]: любые два автоморфизма Бернулли с равной и конечной энтролией слабо изоморфны.

Указанный только что результат получен как следствие более общего утверждения. Именно, пусть у эргодического автоморфизма T имеется возрастающая подалгебра $\mathfrak{M}$. Возьмем любое распределение вероятностей $\{q_i\}$ и

$$-\sum_i q_i \log q_i \leqslant H(T\mathfrak{M}\,|\,\mathfrak{M}).$$

Тогда можно найти такое разбиение пространства M на множества A_1, A_2, ..., A_n, что $A_i \in \mathfrak{M}$, $i=1$, ..., n и для этих множеств A_i при любом целом $k>0$

$$\mu(A_{i_0} \cap T A_{i_1} \cap \ldots \cap T^k A_{i_k}) = q_{i_0} \ldots q_{i_k}$$

Иными словами, автоморфизм с данной энтропией может быть при указанных ограничениях представлен как фактор-автоморфизм любого эргодического автоморфизма с той же энтропией.

Метод доказательства этих утверждений основан на изучении связи между различными возрастающими подалгебрами, соответствующими данному автоморфизму. По существу, уже после первой работы А. Н. Колмогорова [26] стало ясно, что проблема изоморфизма тесно связана с проблемой описания различных возрастающих подалгебр, соответствующих данному автоморфизму. Излагаемый далее метод получения разных возрастающих подалгебр, если известна одна из них, принадлежит В. А. Рохлину. Пусть $\mathfrak{M}$ — возрастающая подалгебра. Тогда всякая подалгебра $\mathfrak{M}_1$, для которой выполняются включения $T^{-1}\mathfrak{M} \subseteq \mathfrak{M}_1 \subseteq \mathfrak{M}$ также будет, как легко видеть, возрастающей. Можно, далее, получать возрастающие подалгебры $\mathfrak{M}_n$ при помощи цепочек переходов:

$$\mathfrak{M}_n \subseteq \mathfrak{M}_{n-1} \subseteq \ldots \subseteq \mathfrak{M}_1 \subseteq \mathfrak{M}, \quad \text{где} \quad T^{-1}\mathfrak{M}_i \subseteq \mathfrak{M}_{i+1} \subseteq \mathfrak{M}_i.$$

Можно, наконец, образовать пересечения

$$\mathfrak{M}' = \bigwedge_i \mathfrak{M}_i, \quad \text{где} \quad T^{-1}\mathfrak{M}_i \subseteq \mathfrak{M}_{i+1} \subseteq \mathfrak{M}_i.$$

Таким образом можно получать новые возрастающие подалгебры $\mathfrak{M}' \subset \mathfrak{M}$. В работе [50] приведены некоторые результаты исследования возможностей, заложенных в таких пересечениях.

Так, оказалось, что если $\mathfrak{M}$ — совершенная подалгебра, то всякая совершенная подалгебра $\mathfrak{M}' \subseteq \mathfrak{M}$ может быть представлена в виде

$$\mathfrak{M}' = \bigwedge_i \mathfrak{M}_i,$$

где $T^{-1}\mathfrak{M}_i \subseteq \mathfrak{M}_{i+1} \subseteq \mathfrak{M}_i$. Более того, чтобы, вообще, пересечение убы-

вающей последовательности совершенных подалгебр было нетривиальным, цепочка переходов от $\mathfrak{M}_i$ к $\mathfrak{M}_{i+1}$ должна быть лишь в известном смысле целенаправленной (разбиения $\eta(\zeta)$ работы [50]).

Наглядный смысл перехода от $\mathfrak{M}$ к $\mathfrak{M}_1$ с включениями $T^{-1}\mathfrak{M} \subseteq \mathfrak{M}_1 \subseteq \mathfrak{M}$ легче уяснить, перейдя от σ-алгебр к соответствующим им измеримым разбиениям. Если ζ и $T^{-1}\zeta$ — разбиения, соответствующие подалгебрам $\mathfrak{M}$ и $T^{-1}\mathfrak{M}$, то каждый элемент $C(T^{-1}\zeta)$ разбиения $T^{-1}\zeta$ состоит из некоторого числа элементов разбиения ζ. Подалгебра $\mathfrak{M}_1$ соответствует разбиению ζ_1 с $T^{-1}\zeta \leqslant \zeta_1 \leqslant \zeta$. Это значит, что элементы $C(\zeta_1)$ разбиения ζ_1 получаются путем «склеивания» элементов $C(\zeta)$ внутри каждого элемента $C(T^{-1}\zeta)$. Доказательство сформулированных выше утверждений о слабом изоморфизме проводится при помощи надлежащей последовательности склеиваний. Интересно, что энтропия выступает здесь не как характеристика неопределенности системы, а как средний коэффициент сжатия. Такой смысл энтропия приобрела в классических динамических системах (см. следующий §).

Представляется весьма правдоподобным, что энтропия составляет полную систему инвариантов слабого изоморфизма для гораздо более широкого класса K-автоморфизмов. В то же время вопрос о полном изоморфизме остается попрежнему открытым и требует, скорее всего, для своего решения новых инвариантов.

§ 3. Энтропия и классические системы

Выяснение разнообразных связей эргодической теории с многочисленными разделами математики и ее внутренние проблемы несколько заслонили основную цель, которая ставилась ее создателями Больцманом, Гиббсом, Пуанкаре, Биргофом. Мы имеем здесь в виду применение методов теории инвариантной меры для исследования поведения решений систем дифференциальных уравнений. В своем известном докладе на Международном Математическом Конгрессе 1954 г. [25] А. Н. Колмогоров дал анализ возможностей, заложенных в таком подходе, подчеркнув чисто метрическую сторону дела. В настоящий момент мы еще не имеем определенных ответов на целый ряд проблем, поставленных в [25]. Однако теперь появилось большое количество новых примеров, иллюстрирующих содержание доклада [25]. Часть этих примеров смогла возникнуть исключительно благодаря энтропии.

Качественное исследование классических динамических систем с интегральным инвариантом развивается в настоящее время по двум направлениям. Это, во-первых, аналитическое исследование окрестностей интегрируемых задач. А. Н. Колмогоровым был введен для этого [24] соответствующим образом переработанный метод Ньютона. Применение этого метода позволило указать [24] целый класс задач, в которых существует множество положительной меры, распадающееся на инвариантные торы с условно-периодическим движением на каждом из них. Последующие исследования В. И. Арнольда [8] показали, что такими, как правило, будут механические системы с двумя степенями свободы, в которых материальная точка движется по выпуклой поверхности.

Другое направление — это исследование динамических систем с положительной энтропией. В основе этих исследований лежит нахожде-

ние условий, при которых система имеет положительную энтропию или является K-системой. Сейчас уже можно считать понятым, что положительность энтропии и наличие перемешивания связаны с крайней неустойчивостью движения системы: траектории выходящие из близких точек, должны разбегаться, вообще говоря, с экспоненциальной скоростью. Следствием такого разбегания является существование, наоборот, асимптотических траекторий, т. е. траекторий, сближающихся с экспоненциальной скоростью. Асимптотические траектории употреблялись и ранее в исследованиях Э. Хопфа [59] и Хедлунда [58], посвященных перемешиванию геодезических потоков на поверхностях отрицательной кривизны. Но их роль стала по-настоящему ясна только в связи с энтропией.

Формализованное изложение связи между существованием асимптотических траекторий и положительностью энтропии основано на понятии трансверсального потока. Мы предполагаем теперь с целью избежать лишних теоретико-множественных трудностей, что пространство с мерой M есть компактное риманово многообразие класса C^∞ (возможно с краем), а инвариантная мера задается положительной плотностью по мере, индуцированной римановой метрикой. Поток $\{S_t\}$ определяется векторным полем на M, компоненты которого имеют ограниченные производные первых двух порядков, в случае же отдельного автоморфизма T функции, задающие преобразование, дифференцируемы и их производные ограничены.

Однопараметрическая группа $[Z_u\}$ преобразований пространства M (не обязательно сохраняющих меру μ!), задаваемая системой дифференциальных уравнений:

$$\frac{dx}{du}=\chi(x) \quad (x\in M), \tag{3}$$

где векторное поле $\chi(x)$ имеет ограниченные производные, является трансверсальным потоком к автоморфизму T, (к потоку $\{S_t\}$), если

1) траектории системы (3) переходят под действием автоморфизма T (каждого автоморфизма S_τ, входящего в поток $\{S_t\}$) в траектории системы (5);

2) для функции $w(u,x)(w_\tau(u,x))$, равной временной длине отрезка $\{T Z_v x,\ 0\leqslant v\leqslant u\}$ $(\{S_\tau Z_v X,\ 0\leqslant v\leqslant u\})$ (u-время) траектории трансверсального потока $\{Z_u\}$, существует

в случае автоморфизма $\lim\limits_{u\to 0}\dfrac{W(u,x)}{u}=\lambda(x)$

в случае потока $\lim\limits_{\tau\to 0}\lim\limits_{u\to 0}\dfrac{W_\tau(u,x)-u}{\tau\cdot u}=\alpha(x).$

Если траектории динамической системы $\{S_t\}$ являются экстремалями некоторой вариационной задачи и имеется поле этих экстремалей, то известная теорема Гаусса из вариационного исчисления [67], [68] показывает, что поле трансверсалей к этому полю экстремалей удовлетворяет условию 1). Этим обстоятельством и объясняется термин «транс-

версальный поток». В вариационном исчислении связь между полем экстремалей и полем трансверсалей изучается обычно докально. Условие 1) — это требование существования поля трансверсалей в целом. Функции $\lambda(x)$ и $\alpha(x)$, фигурирующие в условии 2, естественно назвать коэффициентами сжатия.

Связь трансверсальных потоков с возрастающими подалгебрами видна из следующих двух утверждений.

А. Если автоморфизм T эргодичен и обладает трансверсальным потоком $\{Z_u\}$, для которого $\int_M \log \lambda(x) d\mu < 0$, то существует возрастающая подалгебра $\mathfrak{M}$, такая, что

$$\bigvee_k T^k \mathfrak{M} = \mathfrak{S}, \quad \bigwedge_k T^k \mathfrak{M} = \mathfrak{N}_Z,$$

где $\mathfrak{N}_Z$ σ-алгебра инвариантных mod 0 относительно потока $\{Z_u\}$ множеств. В частности, если трансверсальный поток эргодичен, т. е. $\mathfrak{N}_Z = \mathfrak{N}$, то T есть K-автоморфизм; если $\lambda(x) < 1$ всюду, то требование эргодичности автоморфизма T можно опустить.

В. Если поток $\{S_t\}$ эргодичен и имеет трансверсальный поток $\{Z_u\}$, для которого $\int \alpha(x) d\mu < 0$, то существует возрастающая относительно потока $\{S_t\}$ подалгебра $\mathfrak{M}$ такая, что

$$\bigvee_t S_t \mathfrak{M} = \mathfrak{S}, \quad \bigwedge_t S_t \mathfrak{M} = \mathfrak{N}_Z$$

(логарифм двоичный). Если $\alpha(x) < 0$ почти всюду, то требование эргодичности потока $\{S_t\}$ можно отбросить.

Подалгебра $\mathfrak{M}$ в обоих случаях возникает из специально строящегося разбиения пространства M на отрезки траекторий трансверсального потока. Если $\int \log \lambda(x) d\mu > 0$ или $\int \alpha(x) d\mu > 0$ то утверждение теоремы будет выполняться для автоморфизма T^{-1} или потока $\{S_{-t}\}$.

Наиболее часто встречается случай автоморфизмов, когда $\lambda(x) < 1$ всюду. Тогда для двух точек x_1 и x_2, лежащих на одной траектории трансверсального потока, расстояние между точками $T^n x_1$ и $T^n x_2$ стремится к нулю при $n \to \infty$ с экспоненциальной скоростью, т. е. точки x_1 и x_2 описывают асимптотические траектории. Таким образом, разбиение пространства на траектории трансверсального потока есть разбиение на классы асимптотических траекторий исходной системы.

Условная энтропия подалгебры $\mathfrak{M}$ пропорциональна в случае автоморфизма, среднему логарифму коэффициента сжатия. Таким образом, энтропия характеризует здесь скорость сближения асимптотических траекторий или, что то же, скорость их разбегания. Хотя энтропия есть характеристика движения системы в целом, мы видим, что она выражается через среднее некоторой функции точки. Примеры показывают, что эта функция точки зависит на самом деле от всей траектории данной точки (см. следующий параграф).

Приведем здесь два примера применения трансверсальных потоков. Если T — групповой автоморфизм двумерного тора, то в качестве трансверсального потока $\{Z_u\}$ можно взять движение с единичной скоростью в направлении одного из собственных векторов матрицы, задающей автоморфизм T. Коэффициент $\lambda(x)$ в этом случае постоянен и равен

соответствующему собственному значению. Нетрудно показать, что трансверсальный поток эргодичен в случае эргодического автоморфизма T. Из А. следует, что такой автоморфизм тора является K-автоморфизмом. Другой пример — геодезический поток на поверхности постоянной отрицательной кривизны. В этом случае трансверсальным потоком является хорошо известный поток орициклов (см. [59], [58], [45]). Эргодичность потока орициклов была установлена Хедлундом [58]. Здесь коэффициент $\alpha(x)=\sqrt{-K}$, где K — кривизна поверхности. Таким образом, геодезический поток на поверхности постоянной отрицательной кривизны является K-системой.

Для того, чтобы охватить динамические системы в пространствах числа измерений, большего трех, нужно расширить понятие трансверсального потока до трансверсального поля. Трансверсальным полем естественно считать расслоение пространства M на траектории поля, где отдельная траектория представляет собой многообразие размерности, большей 1, причем это расслоение должно обладать по отношению к автоморфизму T или потоку $\{S_t\}$ свойством инвариантности, сформулированным в условии 1, требуемом от трансверсального потока. Что касается условия 2, то теперь естественно требовать существование коэффициента сжатия вдоль каждого направления в касательной плоскости к траектории трансверсального поля. За параметр u можно принять длину дуги. Основной результат сохраняется в данном случае при предположении, что коэффициент сжатия по каждому направлению при всех для автоморфизма меньше 1 и отрицателен для потока. Расслоение на траектории трансверсального поля снова будет в этом случае расслоением на асимптотические траектории. В применении к геодезическим потоком на многообразиях постоянной отрицательной кривизны это соответствует переходу от потока орициклов к полю орисфер.

Интересно отметить, что орициклы и орисферы появились в работе И. М. Гельфанда и М. И. Граева [16] в связи с разложением некоторых представлений групп Ли на неприводимые компоненты. В рассматриваемых вопросах, вообще, проблемы эргодической теории и проблемы теории представлений чрезвычайно близки. Это стало ясным уже после работы И. М. Гельфанда и С. В. Фомина [19]. В настоящее время происходит известная конкуренция в применении к задачам эргодической теории алгебраических методов и чисто метрических. Метрические методы применимы в более общей обстановке, но алгебраические методы там, где они применимы, позволяют получать несколько более детальные результаты (см. следующий §).

§ 4. Геодезические потоки и другие динамические системы в пространстве линейных элементов поверхности

Механические системы можно во многих случаях свести к геодезическим потокам. В настоящее время наиболее подробно исследованы системы с двумя степенями свободы, сводящиеся к геодезическим потокам на двумерных поверхностях. Сказанное выше об экспоненциаль-

ной быстроте разбегания близких траекторий следует теперь относить к геодезическим, выходящим из одной точки.

Если поверхность выпукла, то уже приводившийся результат В. И. Арнольда [8], показывает, что у такой поверхности существует, вообще говоря, множество линейных элементов положительной меры, распадающиеся на инвариантные торы с дискретным спектром. Вряд ли энтропия может быть полезна для этого случая. Пусть теперь поверхность есть тор. Из исследований Хедлунда [55], Хедлунда и Морса [57], Э. Хопфа [61] и Грина [54] вытекает естественное предположение, что в этом случае скорость разбегания геодезических не более, чем степенная. Это предположение не доказано в метрических терминах, но есть все основания предпологать, что и в этом случае энтропия геодезического потока равна нулю.

Перейдем к поверхностям рода большего единицы. Среди таких поверхностей имеются поверхности постоянной отрицательной кривизны, для которых выше было установлено, что соответствующий геодезический поток является K-системой. Здесь естественно ставить вопрос при каких римановых метризациях поверхности геодезический поток будет K-системой. Примыкая к предыдущему параграфу, исследуем в этом случае проблему существования трансверсального потока. Первое условие, которое следует наложить, это — отсутствие у соответствующей вариационной задачи сопряженных точек. Ясно, что оно выполняется у геодезического потока на поверхности отрицательной кривизны. Отсутствие сопряженных точек выполняется даже при существовании областей положительной кривизны, но при этом всякая геодезическая должна, во-первых, находиться не слишком долго в области положительной кривизны, и, во-вторых, после пребывания в области положительной кривизны она должна провести достаточно большое время в области отрицательной кривизны (см. Э. Хопф [60]). С геометрической стороны отсутствие сопряженных точек означает, что геодезические, выходящие из одной точки под близкими углами, прежде чем пересечься, должны далеко друг от друга разойтись. К настоящему моменту доказано, что отсутствие сопряженных точек обеспечивает лишь положительность энтропии геодезического потока. Пусть q — произвольная точка поверхности и Γ — проходящая через нее геодезическая. В качестве параметра на Γ возьмем длину t, отсчитываемую от точки q. Тогда отсутствие сопряженных точек означает, что решение вариационного уравнения Якоби

$$y'' + K(t)\,y = 0,$$

где $K(t)$ — гауссова кривизна поверхности в точке t, удовлетворящее начальным условиям $y(t_0)=0$, $y'(t_0)=1$, не обращается в ноль при $t>0$. Посредством замены $x(t)=y'(t)/y(t)$ перейдем к соответствующему уравнению Риккати

$$x' + x^2(t) + K(t) = 0. \tag{4}$$

Решение этого уравнения в точке t с начальным условием $x(t_0)=\infty$ имеет простой геометрический смысл — это геодезическая кривизна геодезической окружности радиуса $t-t_0$ с центром в t_0 в точке пересе-

чения этой окружности с геодезической Γ. Зафиксировав t, устремим t_0 к $-\infty$. Оказывается, что для почти всех геодезических соответствующие решения $\varkappa(t)$ сходятся к некоторому всюду конечному положительному решению уравнения (4). Естественно считать, что это решение дает геодезическую кривизну орицикла, проходящего через t ортогонально к Γ. Пока что удалось показать, что энтропия геодезического потока пропорциональна интегралу по пространству от этого решения. С другой стороны, наложив несколько более сильное требование «неустойчивости» по Хопфу [60], можно доказать, что геодезический поток будет K-системой. Эргодичность в этом случае была доказана Хопфом [60] давно. В класс поверхностей, удовлетворяющих условию «неустойчивости» Хопфа, входят поверхности отрицательной кривизны.

В пространстве линейных элементов поверхности отрицательной кривизны можно рассмотреть, следуя В. И. Арнольду [7], другие динамические системы. Зададим некоторое число ϱ и рассмотрим на поверхности линии Γ_ϱ постоянной геодезической кривизны ϱ. Каждый линейный элемент определяет одну из линий Γ_ϱ, которой он касается. Пусть поток $\{S_t\}$ состоит в перемещении с единичной скоростью линейного элемента вдоль определяемой им линии Γ_ϱ. В частности, при $\varrho=0$ получаем геодезический поток. Оказывается, что этот поток сохраняет ту же меру, что и геодезический поток [7]. Свойства этого потока зависят отвеличины ϱ. Предположим, что поверхность имеет постоянную отрицательную кривизну K. Тогда при $0\leqslant\varrho<\sqrt{-K}$ траектории потока $\{S_t\}$ будут эквидистантами и все эти потоки изоморфны геодезическому потоку. Поток орициклов ($\varrho=\sqrt{-K}$) является трансверсальным ко всем этим потокам. При $0\leqslant\varrho<\sqrt{-K}$ все потоки $\{S_t\}$ являются K-системами, имеют счетнократный лебеговский спектр, перемешивание всех степеней и т. д. При $\varrho=\sqrt{-K}$ поток $\{S_t\}$ есть, как уже говорилось, поток орициклов. Б. М. Гуревич [21] показал, что энтропия потока орициклов равна нулю. Следовательно, поток орициклов уже не может быть K-системой. Однако, он эргодичен и имеет даже счетнократный лебеговский спектр (О. С. Парасюк [31]). Неизвестно, обладает ли он перемешиванием высших степеней. При $\varrho>\sqrt{-K}$ характер потока $\{S_t\}$ резко меняется. Теперь его траектории делаются замкнутыми линиями и он становится неэргодическим, имея на эргодических компонентах дискретный спектр. Поток орициклов, таким образом, отделяет друг от друга системы разных типов. Мы привели здесь этот пример для иллюстрации того, как при изменении параметров системы происходит изменение характера ее движения с переходом от случая K-системы к случаю дискретного спектра.

Так как геодезические потоки на многообразиях отрицательной кривизны являются K-системами, т. е. движение любого множества A делается через некоторое время статистически независимым от положения любого другого множества B, то для всякой функции $f(x)$ интеграл

$$\int_0^T f(S_t x)\,dt$$

складывается из слабо зависимых между собой слагаемых. Естественно, поэтому ожидать, что при известных условиях нормированная величина

$$\frac{1}{\sqrt{D_T(f)}}\left[\int_0^T f(S_t x)\,dt - T\bar{f}\right],$$

где
$$\bar{f}=\int_M f(x)\,d\mu, \quad D_T(f)=\int_M\left[\int_0^T f(S_t x)\,dt - T\bar{f}\right]^2 d\mu,$$

будем иметь при $T\to\infty$ гауссовское распределение. Известные в теории вероятностей методы доказательства центральной предельной теоремы для слабо зависимых стационарных рядов предполагают выполненным условие «сильного перемешивания» М. Розенблатта (см. [64], [15], [22]).

Для геодезического потока не удается установить, что естественного вида функции удовлетворяют условию Розенблатта. Однако, можно ввести так называемое «локальное условие Розенблатта» (см. [49]) и модифицировать доказательство центральной предельной теоремы так, что оно станет пригодным для достаточно широкого класса функций $f(x)$ на пространстве линейных элементов поверхности отрицательной кривизны и геодезического потока $\{S_t\}$. Удается показать [46], что в случае геодезического потока $\{S_t\}$ в пространстве линейных элементов на поверхности отрицательной кривизны класс функций $f(x)$, для которых при любом α, $-\infty<\alpha<\infty$

$$\lim_{T\to\infty}\mu\left\{x:\frac{\int_0^T f(S_t x)\,dt - T\bar{f}}{\sqrt{D_T(f)}}<\alpha\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha}\exp\left(-\frac{u^2}{2}\right)du.$$

будет всюду плотен в $\mathfrak{L}^2$. В случае поверхности постоянной отрицательной кривизны удается пойти несколько дальше (см. [46]) по пути конкретного изучения запаса функций, подчиняющихся центральной предельной теореме. Здесь приходит на помощь теория представлений, которая позволяет доказать, что среди таких функций существуют функции, имеющие реальное геометрическое содержание. Например, можно взять в качестве функции характеристическую функцию множества линейных элементов, носители которых лежат внутри некоторой области на поверхности. Тогда оказывается, что нормированное время пребывания геодезической в этой области имеет гауссовское распределение. В [46] имеются и другие примеры. Центральную предельную теорему можно доказать и для геодезического потока в многомерном пространстве постоянной отрицательной кривизны.

Центральная предельная теорема указывает на совсем запутанный характер отдельной траектории геодезического потока. Здесь мы видим пример новой и довольно неожиданной связи между теорией вероятностей и классическими динамическими системами. Вполне возможно, что в общем случае в реальных физических задачах, где понятие отдельной траектории вообще не имеет смысла, имеются K-системы, перемешивание, центральная предельная теорема и т. д. Некоторые перспективы приложения энтропийной теории к статистической механике видны уже сейчас (см. следующий §).

§ 5. Примеры

Может создаться впечатление, что рассмотренные в предыдущем § K-системы имеют характер специально построенных примеров и что для реальных механических задач вся теория не применима. По этому поводу следует сразу же напомнить, как еще в докладе А. Н. Колмогорова [25] указывалось, что можно около поверхности рода, большего единицы, расположить некоторое число центров притяжения и отталкивания и, заставив материальную точку двигаться по поверхности, получить динамическую систему, сводящуюся к геодезическому потоку на поверхности отрицательной кривизны. Совсем простой пример динамической системы с положительной энергией получается следующим образом (доказанная мною гипотеза В. И. Арнольда). Внутри двумерного тора T^2, реализованного как единичный квадрат с попарно отождествленными сторонами рассматривается замкнутая выпуклая гладкая кривая. Материальная точка движется по тору вне этой кривой. Достигая последней, она отражается от нее по закону «угол падения равен углу отражения». Эта система имеет положительную энтропию, трансверсальный поток; траектории, выходящие из одной точки, разбегаются с экспоненциальной быстротой и т. д. Эргодичность в этом примере пока не установлена. Причина положительности энтропии состоит здесь в том, что описанную динамическую систему можно свести к геодезическому потоку по некоторой вырожденной поверхности с краем, у которой кривизна равна всюду нуля, кроме края, где ее естественно считать равной $-\infty$.

Еще один совсем конкретный пример получается из системы двух твердых плоских дисков, двигающихся по инерции на торе и сталкивающихся по законам упругого удара. Эта динамическая система неэргодична на многообразии постоянной энергии (помимо закона сохранения энергии есть закон сохранения импульса) и на почти каждой эргодической компоненте имеет положительную энтропию.

Ограничение плоским случаем здесь несущественно и результат о положительности энтропии верен и для системы двух твердых шариков, двигающихся внутри трехмерного тора. Интересная особенность этих случаев состоит в том, что траектории трансверсального потока здесь оказываются только кусочно-гладкими и не подходят формально под определение § 3. Систему, из двух дисков или двух шариков можно представить как систему двух материальных точек, взаимодействующих друг с другом (отталкивающихся) только в момент, когда расстояние между ними станет равным диаметру диска или шарика. Было бы интересным выделить вообще все возможные виды короткодействующих потенциалов взаимодействия двух точек, определяющих эргодическую систему с положительной энтропией или даже K-систему.

Далее, простейшая механическая модель идеального газа — это система твердых шариков (молекул), движущихся внутри прямоугольного ящика и сталкивающихся по законам упругого удара. Повидимому, это тоже есть пример системы с положительной энтропией и с трансверсальным потоком. Заметим в этой связи, что физиками уже давно выдвигалась идея о необходимости в статистической механике систем с перемешиванием. В первую очередь здесь следует упомянуть Н. С. Крылова [27]. Он, повидимому, первым обратил внимание на то,

что в системах статистической механики есть экспоненциальность разбегания траекторий.

В [27] имеются даже величины, близкие к энтропии. Там они связываются со временами релаксации. Вполне возможно, что энтропийная теория поможет обосновать эргодическую гипотезу в статистической механике, хотя сомнительно, чтобы современная физика получила от этого какую-либо пользу. Но энтропийные методы могут, безусловно, быть полезными при качественном исследовании поведения систем статистической механики.

Заключение

Выбор материала связан обычно с интересами автора и потому всегда в той или иной степени субъективен. Хотелось бы только упомянуть об очень интересных работах А. М. Вершика о динамических системах, порожденных гауссовскими стационарными процессами [12], [13], а также исследования Грина [53], Грина, Маркуса, Ауслендера и Хана [52], [56] так называемых нильпотоков, которые естественно связать с исследованиями Л. М. Абрамова [4] автоморфизмов с квазидискретным спектром.

Цель настоящего доклада можно считать достигнутой, если удалось убедить слушателей, что мы переживаем в настоящее время новый период расцвета эргодической теории. Хотелось бы надеяться, что ближайшие годы будут годами дальнейших успехов этой науки.

В заключение приношу благодарность Организационному Комитету Конгресса за приглашение выступить здесь с этим докладом.

БИБЛИОГРАФИЯ(1)

[1]. Абрамов, Л. М., Энтропия автоморфизма соленоидальной группы. *Теория вероятн. и ее примен.*, 4, 3 (1959), 249–254.

[2]. —— Энтропия производного автоморфизма. *ДАН СССР*, 128, 4 (1959), 980–983.

[3]. —— Об энтропии потока. *ДАН*, 128, 5 (1959), 873–876.

[4]. —— Автоморфизмы с квазидискретным спектром. *Изв. АН СССР, сер. матем.*, 27, 4 (1962).

[5]. Абрамов, Л. М., и Рохлин, В. А., Энтропия косого произведения преобразований с инвариантной мерой. *Вестник Ленинградского Университета*, 2, 7 (1962), 5–13.

[6]. Аносов, Д. В., Грубость геодезических потоков на компактных многообразиях отрицательной кривизны. *ДАН СССР*, 145, 4 (1962).

[7]. Арнольд, В. И., Несколько замечаниях о потоках линейных элементов и реперов. *ДАН СССР*, 138, 2 (1961), 255–257.

[8]. —— Об устойчивости положения равновесия гамильтоновой системы обыкновенных дифференциальных уравнений в общем эллиптическом случае. *ДАН СССР*, 137, 2 (1961), 255–257.

[9]. —— Замечания о числах вращения. *Сибирский математ. журнал* 2, 6 (1961), 807–813.

(1) В библиографию включены работы последних лет по эргодической теории или связанные с содержанием доклада.

[10]. Арнольд, В. И., и Синай, Я. Г., О малых возмущениях автоморфизмов тора. *ДАН СССР*, 144, 4 (1962), 695–698.

[11]. Беляев, Ю. К., Один пример процесса с перемешиванием. *Теория вероятностей и ее прим.*, IV в.1 (1961), 101–103.

[12]. Вершик, А. М., К теории нормальных динамических систем. *ДАН СССР*, 144, 1(1962), 9–12.

[13]. —— О спектральном и метрическом изоморфизме некоторых нормальных динамических систем. *ДАН СССР*, 144, 2 (1962), 255–257.

[14]. Виноградов, В. Г., Условия регулярности вероятностных процессов. *ДАН СССР*, 113, 5 (1957), 959–961.

[15]. Волконский, В. А., и Розанов, Ю. А., Некоторые предельные теоремы для случайной функции I. *Теория вероятностеи и ее прим.*, 4, 2 (1959).

[16]. Гельфанд, И. М., и Граев, М, И., Геометрия однородных пространств, представления групп в однородных пространствах и связанные с ними вопросы интегральной геометрии. I. *Труды Москов. Математ. Общества*, 8 (1959), 321–390.

[17]. Гельфанд, И. М., и Пятецкий–Шапиро, И. И., Об одной теореме Пуанкаре. *ДАН СССР*, 127, 3 (1959), 490–493.

[18]. —— Теория представлений и теория автоморфных функций. *УМН*, 14 в.2 (1959), 171–194.

[19]. Гельфанд, И. М., и Фомин, С. В., Геодезические потоки на многообразиях постоянной отрицательной кривизны. *УМН*, 7 в.1 (1952), 118–137.

[20]. Гирсанов, И. В., О спектрах динамических систем, порождаемых стационарными гауссовскими процессами. *ДАН*, 119, 5, 851–853.

[21]. Гуревич, Б. М., Энтропия потоков орициклов. *ДАН СССР*, 136, 4 (1961).

[22]. Ибрагимов, И. А., Некоторые предельные теоремы для стационарных в узком смысле вероятностных процессов. *ДАН СССР*, 125, 4 (1959).

[23]. —— Одна теорема из метрической теории цепных дробей. *Вестник Ленинград. Университета*, 1, 1 (1961) 13–24.

[24]. Колмогоров, А. Н., О сохранении условно-периодических движений при малом изменении функции Гамильтона. *ДАН*, 98, 4 (1954), 527–530.

[25]. —— Общая теория динамических систем и классическая механика. *Proc. Internat Congress Math. Amsterdam*, 1 (1954), 315–333.

[26]. —— Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега. *ДАН СССР*, 119, 5 (1958), 861–868.

Об энтропии на единицу времени как метрическом инварианте автоморфизмов. *ДАН СССР*, 124, 4 (1959), 754–755.

[27]. Крылов, Н. С., Работы по обоснованию статистической физики. *Изд. АН СССР*, 1950 г.

[28]. Леонов, В. П., О дисперсии временных средних стационарного случайного процесса. *Теория вероятностей и ее прим.*, VI, 1 (1961), 93–101.

[29]. —— О центральной предельной теореме для эргодических эндоморфизмов компактных коммутативных групп. *ДАН*, 135, 2 (1960), 258–261.

[30]. Мешалкин, Л. Д., Один случай изоморфизма схем Бернулли. *ДАН СССР*, 128, 1 (1959), 41–44.

[31]. Парасюк, О. С., Потоки орициклов на поверхностях постоянной отрицательной кривизны. *УМН*, 8, 3 (1953), 125–126.

[32]. Пинскер, М. С., Динамические системы с нулевой и вполне положительной энтропией. *ДАН СССР*, 133, 5 (1960), 1025–1026.

[33]. —— Информация и информационная устойчивость случайных величин и процессов. Изд. *АН СССР*, 1960.

[34]. Рохлин, В. А., Об основных понятиях теории меры. *Матем. сборник*, 25 (67): 1 (1949), 107–150.

[35]. —— Избранные вопросы метрической теории динамических систем. *УМН*, IV, 2 (1949), 57–128.

[36]. Об эндоморфизмах коммутативных компактных групп. *Изв. АН СССР, сер. матем.*, 13 (1949), 329–340.

[37]. —— Об энтропии метрического автоморфизма. *ДАН*, 124, 5 (1959), 980–983.

[38]. —— Новый прогресс в теории преобразований с инвариантной мерой., *УМН*, XV, 4 (1960), 3–26.

[39]. —— Точные эндоморфизмы пространства Лебега. *Изв. АН СССР, сер. матем*, 25 (1961), 499–530.

[40]. —— Об энтропии автоморфизма компактной коммутативной *группы. Теория вероятн. и ее прим.*, 5, 3 (1961), 351–352.

[41]. Рохлин, В. А., и Синай, Я. Г., Построение и свойства инвариантных измеримых разбиений. *ДАН СССР*, 141, 5 (1961), 1038–1041.

[42]. Синай, Я. Г., О понятии энтропии динамической системы. *ДАН СССР*, 124, 4 (1959), 768–771.

[43]. —— О потоках с конечной энтропией. *ДАН СССР*, 125, 6 (1959), 1200–1202.

[44]. —— Динамические системы и стационарные марковские процессы *Теория вероятн. и ее прим.*, 5, 3 (1960), 335–338.

[45]. —— Геодезические потоки на многообразиях отрицательной постоянной кривизны. *ДАН СССР*, 131, 4 (1960), 752–755.

[46]. —— Центральная предельная теорема для геодезических потоков на многообразиях постоянной отрицательной кривизны. *ДАН СССР*, 133, 6 (1960), 1303–1306.

[47]. —— Динамические системы со счетнократным лебеговским спектром I. *Изв. АН СССР, сер. матем.*, 25, 6 (1961), 899–924.

[48]. —— Геодезические потоки на компактных поверхностях отрицательной кривизны. *ДАН*, 136, 3 (1961), 549–552.

[49]. —— О предельных теоремах для стационарных процессов. *Теория вероятн. и ее прим.*, 7, 2 (1962), 213–219.

[50]. —— Слабый изоморфизм преобразований с инвариантной мерой. *ДАН СССР* (в печати).

[51]. Хинчин, А. Я., Об основных теоремах теории информации. *УМН*, 11, 1 (1956), 17–77.

[52]. Green, L., Auslander, L. & Hahn, F., Flows on some three dimensional homogeneous spaces. *Bull. Amer. Math. Soc.*, 67, 5 (1961), 494–497.

[53]. Green, L., Spectra of nilflows. *Bull. Amer. Math. Soc.*, 67, 4 (1961). 414–415.

[54]. —— Surfaces without conjugate points. *Trans. Amer. Math. Soc.*, 76, 3 (1954), 529–546.

[55]. Hedlund, G. A., Geodesics on a two-dimensional Riemannian manifold with periodic coefficients. *Ann. Math., Ser. 2*, 33, 4 (1932), 719–739.

[56]. Auslender, L., Hahn, L. & Markus, Topological dynamics. *Bull. Amer. Math. Soc.*, 67, 3 (1961).

[57]. Morse, M., & Hedlund, G., Manifolds without conjugate joints. *Trans. Amer. Math. Soc.* 51, 2 (1942), 362–386.

[58]. Hedlund, G., The dynamics of geodesic flows. *Bull. Amer. Math. Soc.*, 45, 4 (1939), 241–245.

[59]. Хопф, Э., Статистика геодезических линий на многообразиях отрицательной кривизны I. *УМН*, IV, 2 (1949), 129–170.

[60]. Hopf, E., Statistik der Lösungen geodätischer Probleme vom unstabilen Typus II. *Math. Ann.*, 117*b*, 4 (1940), 590–608.

[61]. —— Closed surfaces without conjugate points. *Proc. Nat. Acad. Sci. U.S.A.*, 34 (1948), 47–51.

[62]. Halmos, P., Recent progress in ergodic theory. *Bull. Amer. Math. Soc.*, 67, 1 (1961), 70–80.

[63]. Rosenblatt, M., A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42, 1 (1956), 43.

[64]. —— Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.*, 8, 5 (1959), 665–681.

[65]. —— Stationary Markov chains and independent random variables. *J. Math. Mech.*, 9, 6 (1960), 945–949.

[66]. Бляшке, В., *Основы дифференциальной геометрии.* М.–Л., 1936.

[67]. Лаврентьев, М. А., и Люстерник, Л. А., *Курс вариационного исчисления.* М.–Л. ГТТИ, 1950.

MATHEMATICAL PROBLEMS IN MAGNETO-FLUID DYNAMICS AND PLASMA PHYSICS(1)

By HAROLD GRAD

It is my particular pleasure to have the honor to present an address on the subject of magneto-fluid dynamics in the home of the founders of the subject, Alfvén, Lundquist, Herlofson, and their associates. This is a very new subject, not only in mathematics but even in physics. Generally a young subject in physics is found to follow somewhat pragmatic rules of evidence rather than strict mathematical deduction; historically, it is only at a later stage that we find a structure which is sufficiently formalized to be able to qualify as mathematics. But, because of the very rapid pace of modern science, this subject has moved very quickly into the realm of mathematics. This is not to say that the subject is now ready to be embalmed by Nicholas Bourbaki, but rather that there exists, at least in some aspects, a recognizable mathematical structure. Although I cannot hope to cover the entire field in this lecture, I shall attempt to briefly survey most of the mathematically significant high points, mainly work done in New York University at the Courant Institute. We shall examine, on the one hand, places where familiar analytical tools appear in a new setting and, on the other hand, situations where the mathematical structure is not so familiar but where, if history is a guide, we may expect the growth of new legitimate mathematical enterprises. In particular, we shall find that the classical trichotomy of partial differential equations into elliptic, hyperbolic, and parabolic types leaves uncovered a significant part of the subject. Although even the more unusual equations we encounter have solutions in the most general sense [1], the deeper question of the manifold of *all* solutions (described, as usual, in terms of well-posed initial and boundary value problems) is only beginning to be developed.

1. The equations

The physical problem concerns the flow of an electrically conducting medium in the presence of an electromagnetic field. On the one hand, the field exerts a force on the fluid; on the other hand, the motion generates charge and current sources which reflect back on the electromagnetic field. Perhaps the most striking feature of this interaction is the interplay between the scalar pressure of the fluid and the extremely anisotropic electromagnetic stress.

The equations we use are essentially those introduced by Lundquist [2]:

$$\frac{d\varrho}{dt} + \varrho \operatorname{div} \mathbf{u} = 0, \tag{1.1}$$

(1) The material on Plasma Physics has been omitted from the written manuscript; for a survey of this subject we refer to H. Grad, *Modern Kinetic Theory of Plasmas*, in the Proc. of the Fifth International Conference on Ionization Phenomena in Gases, Munich 1961, North-Holland Publishing Co.

$$\varrho\frac{d\mathbf{u}}{dt}+\nabla p=\mathbf{J}\times\mathbf{B},\quad \mathbf{J}=\frac{1}{\mu_0}\operatorname{curl}\mathbf{B}, \tag{1.2}$$

$$\frac{d\eta}{dt}=0, \tag{1.3}$$

$$\left.\begin{aligned}&\frac{\partial\mathbf{B}}{\partial t}+\operatorname{curl}\mathbf{E}=0\\&\mathbf{E}+\mathbf{u}\times\mathbf{B}=0\end{aligned}\right\}\frac{\partial\mathbf{B}}{\partial t}+\operatorname{curl}(\mathbf{B}\times\mathbf{u})=0, \tag{1.4}$$

$$\operatorname{div}\mathbf{B}=0. \tag{1.5}$$

We use $d/dt=\partial/\partial t+\mathbf{u}\cdot\nabla$ for the Lagrangian derivative following the fluid. These equations are Galilean invariant as a consequence of the particular form taken for the electromagnetic equations and the electromagnetic forces. Specifically, the displacement current and electric component of the force are suppressed in (1.2); this allows $\mathbf{J}$ to be eliminated in favor of $\mathbf{B}$.(1) The statement of perfect conductivity, $\mathbf{E}+\mathbf{u}\times\mathbf{B}=0$, combined with the other Maxwell equation, yields the equation (1.4) for the conservation of magnetic flux. This equation (together with (1.5)), familiar from the theory of vorticity, states that individual magnetic lines are carried with the fluid at the velocity $\mathbf{u}$, and the flux per unit area (strength of a magnetic tube) is constant.

The complete system consists of the conservation of mass (1.1), of momentum (1.2), of energy (1.3), and of flux (1.4) and (1.5), together with an appropriate equation of state relating ϱ, p, and η. Qualitatively, this system describes a magnetic field which is carried with the fluid while reacting back on the motion through the force per volume

$$\mathbf{J}\times\mathbf{B}=\frac{1}{\mu_0}\mathbf{B}\cdot\nabla\mathbf{B}-\nabla(B^2/2\mu_0). \tag{1.6}$$

The mathematical significance of this system of equations lies in the fact that it is symmetric hyperbolic and of conservation form [6]. From the hyperbolicity we have access to familiar existence and uniqueness theorems for initial and boundary value problems; in addition, the concepts of domain of dependence and wave propagation are relevant. From the conservation form, we predict some properties of weak solutions, in particular shock waves. But, despite the general familiarity of the mathematical structure, we shall find the detailed properties to be very strange—principally because of the anisotropy introduced by the magnetic field.

The statement that the system of equations is hyperbolic requires some slight elaboration. The system (1.1)–(1.4) (ignoring (1.5)) can be seen to be symmetric hyperbolic by inspection; (in terms of the variable p rather than ϱ, we replace (1.1) by $dp/dt+a^2\varrho\operatorname{div}\mathbf{u}=0$, where a^2 is the ordinary adiabatic sound speed). If the initial data are taken to satisfy $\operatorname{div}\mathbf{B}=0$, and if a suitable boundary condition is imposed (on a moving boundary receding from

(1) A fully relativistic fluid description taken with the complete Maxwell equations yields a very similar general mathematical structure, cf. [3] and [4]. But the combination of conventional fluid equations with the Lorentz-invariant Maxwell equations yields a very undesirable differential system [5].

the fluid), a subclass of solutions is found satisfying div $\mathbf{B}=0$ and subject to a certain smaller domain of dependence. Alternatively, we can parametrize div $\mathbf{B}=0$ in terms of the two Euler stream functions,

$$\mathbf{B}=\nabla\alpha\times\nabla\beta, \tag{1.7}$$

and replace (1.4) by

$$\left.\begin{aligned}\frac{d\alpha}{dt}&=0,\\ \frac{d\beta}{dt}&=0.\end{aligned}\right\} \tag{1.8}$$

In the variables $(\varrho,\eta,\mathbf{u},\alpha,\beta)$ the system (1.1)–(1.3) and (1.8) is hyperbolic, but it is not recognizably symmetric.

The theory of this hyperbolic system is outlined in Sect. 2. Of course, the statement that the time-dependent system is hyperbolic contains no implication with regard to problems of steady flow or static equilibrium; these will be investigated separately in Sects. 4 and 6 respectively. Each of these general formulations (time-dependent, steady flow, static equilibrium) is followed by a special case which can be analysed in more detail (transverse flow in Sect. 3, aligned flow in Sect. 5, the static free boundary in Sect. 7). These special cases all feature the property that they are exact mathematical analogues of problems in fluid dynamics (this is even true of the general static equilibrium). The transverse flow analogue allows the transfer of a large body of fluid dynamic theory to magnetofluid dynamics; in the aligned flow problem this transfer is limited by singularities in the transformation to the analogue variables; in the general static equilibrium, with the recent development of the magnetic version, the analogue is more appropriately used in the opposite direction! Sect. 7 also contains a complete description of the stability of magnetic free boundary configurations.

2. Characteristics and wave propagation

The qualitative and many quantitative features of a hyperbolic system are elucidated by an examination of the characteristic cones. For the system (1.1)–(1.5), these were computed by Friedrichs [6]. There are two degenerate first order characteristic lines immediately recognizable, viz., the particle paths for η and the lines $x=$ constant for div $\mathbf{B}$. In addition there are three distinct nontrivial cones corresponding to a system of sixth order. In Fig. 1 are shown the characteristic loci, i.e., the traces of the characteristic cones at unit time. The figures are surfaces of revolution about the direction of $\mathbf{B}$. These loci are determined by the ratio of two parameters (local properties of the medium), the adiabatic sound speed

$$a=(\partial p/\partial\varrho)^{1/2} \tag{2.1}$$

and the Alfvén speed

$$A=B/(\mu_0\varrho)^{1/2}. \tag{2.2}$$

The characteristic loci are subject to several interpretations. First, the outer convex locus identifies the domain of dependence. Each of the figures describes the propagation of discontinuities, either of derivatives of any order or of small discontinuities in the variables themselves (weak shocks).

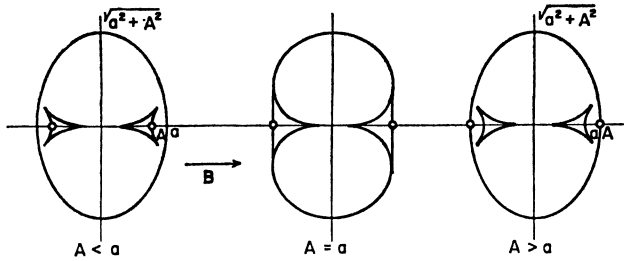


Fig. 1. Characteristic loci.

The various loci in Fig. 1 also separate different analytic pieces of the fundamental solution; the figure is a snapshot of the result of a point explosion. Finally, the various loci in Fig. 1 are to be used instead of spheres (as in an isotropic medium) in the application of Huyghen's construction of "wavelets". We shall also see in Sect. 4 how the characteristics in a steady flow (Mach cones) can be constructed from this figure.

It will be convenient to consider three types of question in order of complexity. First we shall examine the propagataion of plane waves in a linear perturbation of a constant state. This is an essentially algebraic problem. Then we observe the behavior of discontinuities in general solutions or, the equivalent, the propagation of small discontinuous wave fronts superposed on an arbitrary solution. This involves the study of a single first order equation (Hamilton–Jacobi) whose theory reduces to ordinary differential equations. Only then do we examine properties of the full (sixth order) system of partial differential equations.

The propagation of plane waves is completely described by the *normal speed* diagram, Fig. 2. In a given direction are plotted radii proportional to the normal speeds of plane waves with the given orientation. There are three speeds (each with a double multiplicity $\pm c$) in each direction, the *intermediate* speed

$$c^2 = A^2 \cos^2 \theta \tag{2.3}$$

and the *slow* and *fast* speeds (together referred to as *compressive*) given by

$$c^4 - c^2(a^2 + A^2) + a^2 A^2 \cos^2 \theta = 0, \tag{2.4}$$

where θ is the angle between the magnetic field and the plane wave normal. In every direction the slow and fast speeds bracket the intermediate speed. The fast wave is only moderately anisotropic (the maximum ratio of major to minor axis is $2^{\frac{1}{2}}$). Both the intermediate and slow waves are extremely anisotropic and apparently do not propagate at all in the direction perpendicular to **B**. We remark that the characteristic loci, Fig. 1, can be obtained from the normal speed loci, Fig. 2, by taking envelopes of the indicated plane waves; (the compressive loci are given by a certain tenth degree algebraic surface [7]).

Next we turn to the propagation of arbitrary wave fronts (discontinuity surfaces) [8], [9] and examine the characteristic loci. Again we observe that the outer fast (or *magnetosonic*) locus is somewhat anisotropic but is otherwise unexciting. The locus for the intermediate (or *transverse* or *Alfvén*)

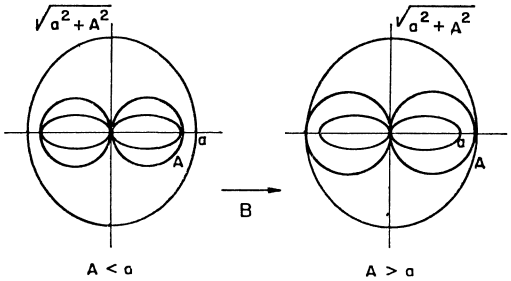


Fig. 2. Normal speed diagram.

wave is extremely degenerate and consists of the two encircled points at $\pm A$ together with the included line segment. This degenerate locus represents strictly one-dimensional propagation along magnetic lines. A transverse wave front is forever contained within a fixed flux tube (but its shape can be distorted in an inhomogeneous medium or a nonuniform field). The slow locus is even more singular. It consists of two disconnected cusped regions together with the intercepted line segment (Fig. 1). The qualitative significance is most easily appreciated by the example, in Fig. 3, of the evolution of an originally spherical slow wavefront in a uniform medium. The slow front lags behind the transverse front (which remains spherical) so long as it is smooth. It overtakes the transverse wave (in apparent contradiction to the implications of the normal speed loci, Fig. 2) after singularities develop in the wave front. Summarizing, the intermediate wave propagates one-dimensionally along flux tubes, the slow wave front is confined to a certain cone surrounding the magnetic field direction, and only the fast wave front propagates in all directions (but not isotropically).

The classical theory of ray optics can be employed to make more quantitative this description of the propagation of wave fronts [8]. A discontinuity or characteristic surface $\phi(x,t)=0$ satisfies the partial differential equation (cf. (2.3) and (2.4))

$$(\phi_t - \mathbf{A}\cdot\nabla\phi)^2[\phi_t^4 - (a^2+A^2)\,\phi_t^2(\nabla\phi)^2 + a^2(\nabla\phi)^2(\mathbf{A}\cdot\nabla\phi)^2] = 0, \qquad (2.5)$$

where $\mathbf{A}$ is the vector Alfvén speed pointing in the direction of $\mathbf{B}$. Each root of (2.5) is a first order partial differential equation which, in turn, has characteristics which can be described as curves rather than cones; these

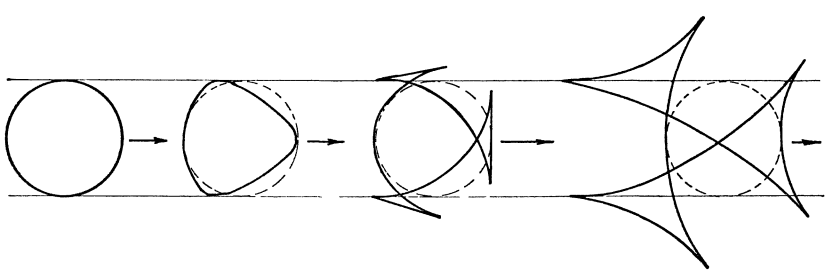


Fig. 3. Slow wave front.

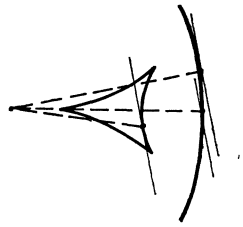


Fig. 4. Ray construction.

are the *rays* or *bicharacteristics*. A given initial discontinuity front splits into six components traveling in both directions along each of the three sets of rays. For each of the categories—slow, intermediate, fast—in a given problem we have a direction field of rays (possibly multiply covering the plane), along which the wave fronts propagate. There is also for each ray an ordinary differential equation governing the variation of the magnitude of the discontinuity front [8]. The singular behavior which results from multi-valued direction fields requires special handling [10].

The ray direction for a given orientation of the wave front is determined as indicated in Fig. 4. For propagation in a homogeneous medium (equation (2.5) has constant coefficients), the ray construction, Fig. 4, is valid in the large. An element of a wave front remains parallel to itself and moves the distance shown along the appropriate straight-line ray; (this remark yields a simple quantitative construction of Fig. 3).

The separation of wave fronts into slow, intermediate, and fast waves is, of course, not possible for general solutions of the original differential system.(1) Even in the case of a linear perturbation about a constant state (constant coefficients), a fast wave front will leave behind it a trail of slow and transverse debris, and a slow wave front will send ahead of it fast and transverse signals (initiated by a wave front of higher order). Nevertheless, in a certain sense the intermediate wave can be partly separated from the compressive wave in some cases [11], [2], [9]. To observe this, consider a linear perturbation about a constant state characterized by parameters ϱ_0 and B_0. The dimensionless perturbation variables

$$\left.\begin{aligned} \sigma &= \varrho/\varrho_0 \\ \beta &= \mathbf{B}\cdot\mathbf{B}_0/B_0^2 \end{aligned}\right\} \tag{2.6}$$

satisfy the fourth order system

$$\left.\begin{aligned} \frac{\partial^2\sigma}{\partial t^2} &= a_0^2\left(\frac{\partial^2\sigma}{\partial x^2}+\frac{\partial^2\sigma}{\partial y^2}+\frac{\partial^2\sigma}{\partial z^2}\right)+A_0^2\left(\frac{\partial^2\beta}{\partial x^2}+\frac{\partial^2\beta}{\partial y^2}+\frac{\partial^2\beta}{\partial z^2}\right) \\ \frac{\partial^2\beta}{\partial t^2} &= a_0^2\left(\frac{\partial^2\sigma}{\partial y^2}+\frac{\partial^2\sigma}{\partial z^2}\right)+A_0^2\left(\frac{\partial^2\beta}{\partial x^2}+\frac{\partial^2\beta}{\partial y^2}+\frac{\partial^2\beta}{\partial z^2}\right) \end{aligned}\right\} \tag{2.7}$$

which has as characteristics the slow and fast cones. On the other hand, the components perpendicular to B_0,

(1) The decomposition is possible at the expense of a loss of domain of dependence attributes, e.g., by a Fourier expansion.

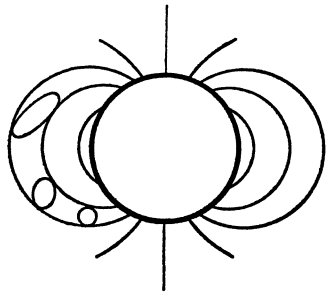


Fig. 5. Whistlers.

$$\left.\begin{aligned} \mathbf{u}' &= \mathbf{u} - \mathbf{B}_0(\mathbf{u}\cdot\mathbf{B}_0)/B_0^2 \\ B_0\boldsymbol{\beta}' &= \mathbf{B} - \mathbf{B}_0\beta \end{aligned}\right\} \tag{2.8}$$

satisfy the inhomogeneous one-dimensional wave equation

$$\left.\begin{aligned} &\frac{\partial \mathbf{u}'}{\partial t} - A_0^2\frac{\partial \boldsymbol{\beta}'}{\partial x} = -A_0^2\nabla'\beta - a_0^2\nabla'\sigma \\ &\frac{\partial \boldsymbol{\beta}'}{\partial t} - \frac{\partial \mathbf{u}'}{\partial x} = 0. \end{aligned}\right\} \tag{2.9}$$

Although (2.9) has the one-dimensional intermediate characteristics, the inhomogeneous term, found from (2.7), is different from zero within the larger compressive domain of dependence.

Despite the fact that an arbitrary initial state cannot be resolved into intermediate and compressive components, each with the appropriate domain of dependence, one can find large classes of initial data that propagate purely as a one-dimensional transverse wave or as a compressive wave. For example, taking $\sigma=\beta=\partial\sigma/\partial t=\partial\beta/\partial t=0$ initially yields a purely transverse solution of (2.9) for $\mathbf{u}'$ and $\mathbf{B}'$ ($\partial\sigma/\partial t=\partial\beta/\partial t=0$ can be satisfied by taking div $\mathbf{u}=0$ and $\mathbf{u}\cdot\mathbf{B}_0=0$ initially). This simple separation does not occur with variable coefficients. As an illustration we consider the phenomenon of "whistlers" in the earth's magnetic field (taking for granted the relevance of these equations). A transverse wave front will propagate along a magnetic tube without radial attenuation and will, under appropriate conditions, be reflected at each end for an indefinite lifetime (Fig. 5). But there is no purely transverse solution of the full system of partial differential equations in a case of variable coefficients. Thus there must be some coupling of the disturbance with compressive waves which radiate away to infinity. The precise relation between the wave front which persists and the complete disturbance which apparently radiates has not yet been established.

An explicit formula has been obtained for the two-dimensional fundamental solution for the case of constant coefficients [12]. In three dimensions, some asymptotic properties of the fundamental solution have been obtained by Fourier analysis [13], [14], but most of these results are more easily obtained using the simpler ray optics techniques.

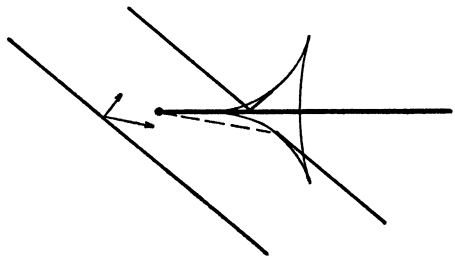


Fig. 6. Wave incident on a slit.

As a final illustration of the unusual wave motions embraced by this theory, we quote an explicit solution for the motion of a plane discontinuity front incident on a semi-infinite slit [15]. The qualitative behavior for a slow wave front is indicated in Fig. 6 (the unperturbed magnetic field is parallel to the slit). Because the ray direction (which is the direction of propagation of the wave) is not normal to the front, we observe that both the shadow and the reflected wave lie on the "wrong" sides of the slit.

3. Transverse flow [16]

There is a simple special configuration which, by appropriate identification of variables, can be made into an exact mathematical analogue with a very large class of ordinary fluid flows. The intriguing feature of this analogue is that it maps a large portion of the entire body of fluid dynamic theory (viz., all two-dimensional nonsteady flows which make no specialization with regard to the equation of state) onto a miniscule corner of magnetofluid dynamics.

Specifically, we assume that all variables depend on (x,y,t), the flow field is two-dimensional (u_x,u_y), and the magnetic field is one-dimensional, having the single component B_z in the ignorable direction. The essential simplification is due to the fact that the Maxwell stress tensor is equivalent to an isotropic scalar pressure in the plane of the flow; cf. (1.6) where $(\mathbf{B}\cdot\nabla)\mathbf{B}=0$. With this remark, the momentum equation

$$\varrho\frac{d\mathbf{u}}{dt}+\nabla p_*=0 \tag{3.1}$$

is formally the same as in fluid dynamics in terms of the "total pressure"

$$p_*=p+B^2/2\mu_0. \tag{3.2}$$

In the flux equation (1.4), we remark that $\mathbf{B}$ can be reinterpreted as a scalar; in this notation (1.4) takes the form

$$\frac{dB}{dt}+B\operatorname{div}\mathbf{u}=0. \tag{3.3}$$

In this geometry, conservation of mass (1.1) is identical to conservation of

flux, since the two-dimensional "volume" element is the same as the element of area. An alternative form to (3.3) is therefore

$$\frac{d\eta_*}{dt}=0, \tag{3.4}$$

where

$$\eta_*=B/\varrho. \tag{3.5}$$

Finally, just as the conventional entropy equation (1.3) is equivalent to an energy equation, (3.4) can be replaced by

$$\varrho\frac{de_*}{dt}+p_* \operatorname{div} \mathbf{u}=0 \tag{3.6}$$

in terms of the "total energy" per mass

$$e_*=e+B^2/2\mu_0\varrho. \tag{3.7}$$

The complete system of equations is the conservation of mass (1.1) and of momentum (3.1) together with the two particle path invariants η and η_*, (1.3) and (3.4).

As it stands, this system is very similar to the ordinary fluid equations (there are the same characteristic cones, with the exception of an extra particle path), and the similarity can be made into an identity by a slight further specialization. We consider three cases:

(1) $\eta_*=$constant,
(2) $\eta=$constant,
(3) $\eta_*=f(\eta)$.

In the first case, we take $\eta_*=B/\varrho$ to be a constant over the entire flow instead of along each particle path. The mass, momentum, and entropy equations are exactly as in fluid dynamics provided that we complete the formulation with an equation of state relating (p_*,ϱ,η), viz.,

$$p_*(\varrho,\eta)=p(\varrho,\eta)+\eta_*^2\varrho^2/2\mu_0. \tag{3.8}$$

It is a simple matter to verify that this equation of state is compatible with conventional thermodynamic convexity requirements, and, in particular, the adiabatic sound speed is given by

$$a_*^2=\partial p_*/\partial\varrho=a^2+A^2. \tag{3.9}$$

In the second case listed above, isentropic flow with η constant, we obtain a conventional system of fluid equations in (p_*,ϱ,η_*), taking η_* as the fluid entropy. Again, we need only an equation of state, in this case

$$p_*(\varrho,\eta_*)=p(\varrho)+\eta_*^2\varrho^2/2\mu_0. \tag{3.10}$$

It is again possible to verify that the thermodynamic structure is standard, and the sound speed takes the same form (3.9).

In the third case, we interpret $\eta_*=f(\eta)$ quite generally to mean that the surfaces $\eta=$constant coincide with $\eta_*=$constant. For example, in one-

dimensional nonsteady flow, variables (x,t), this is no specialization; the relation $\eta_* = f(\eta)$ is obtained by eliminating x between the given initial values of $\eta_*(x)$ and $\eta(x)$. Eliminating η_* in favor of η, we obtain a conventional fluid system with the equation of state

$$p_*(\varrho,\eta) = p(\varrho,\eta) + f^2(\eta)\varrho^2/2\mu_0 \tag{3.11}$$

and the same sound speed (3.9).

The thermodynamic structure in each of the above three special cases can be derived from the elementary identity

$$T\,d\eta + (B/\mu_0)\,d\eta_* = de_* + p_*\,d(1/\varrho). \tag{3.12}$$

In the first example the analogue temperature is $T_* = T$, in the second example it is $T_* = B/\mu_0$, and in the third example it is $T_* = T + Bf'(\eta)/\mu_0$.

The conventional laws of Bernoulli and the conservation of circulation are easily derived. The restricted Bernoulli's law (for a steady flow) is valid in complete generality,

$$h_* + \tfrac{1}{2}u^2 = a(\psi), \tag{3.13}$$

where a is constant on each streamline (as are η and η_*) and the "total enthalpy" is

$$h_* = e_* + p_*/\varrho = h + B^2/\mu_0\varrho. \tag{3.14}$$

Conservation of circulation results under the condition that p_* depend on ϱ alone; this is guaranteed, for example, if both η and η_* are constant throughout the flow. Under the same restrictions potential flow

$$\mathbf{u} = \nabla\phi \tag{3.15}$$

is validated, and we obtain the strong form of Bernoulli's law

$$\frac{\partial\phi}{\partial t} + h_* + \tfrac{1}{2}u^2 = a(t). \tag{3.16}$$

From the analogy with ordinary fluid dynamics we can transfer the relevant theory of shock waves, simple waves and interactions, flows around objects and bends, Riemann invariants, cavitation, etc. But we may not transfer essentially three-dimensional properties related to stability and turbulance, nor dissipative properties connected with boundary layers, etc.

4. Steady flow

The system of equations is as before, (1.1)–(1.5), but this time with the time derivatives suppressed. As a result of this specialization we find the scene dominated by more unusual mathematical structures. A typical problem involves a higher order system which is partly elliptic and partly hyperbolic; e.g. a sixth order system with two real characteristic cones and one imaginary. A more conventional structure which is encountered is a fully hyperbolic system, but purely elliptic problems are not found (except

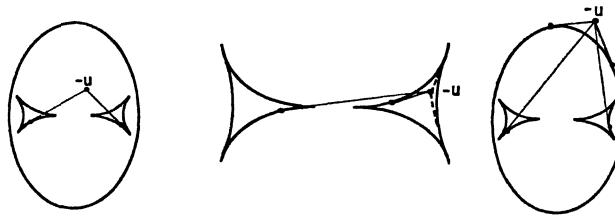


Fig. 7. Steady flow characteristics.

under very special circumstances, see Sect. 5). The classical fluid dichotomy into subsonic (elliptic) and supersonic (hyperbolic) does not generalize. One reason is that the fast and slow characteristics are two real traces of a single complex algebraic surface, and the concepts subsonic and supersonic cannot be unambiguously applied to the slow and fast speeds individually. As to the transverse wave, since propagation is one-dimensional, it is only possible to distinguish subsonic from supersonic in very special (essentially one-dimensional) flows.

The steady-flow characteristic (Mach) cones are obtained from the characteristic loci (Fig. 1) by a simple geometrical construction [16], [17], [18]. With reference to a specific location in a general steady flow, we take the non-steady locus, Fig. 1, with the appropriate local values of a and A, and mark the terminus of the reversed local flow velocity vector $-\mathbf{u}$, on this diagram (Fig. 7). The real characteristic cones are obtained by drawing tangents from $-\mathbf{u}$ to the characteristic loci;([1]) in two dimensions we obtain lines and in three dimensions cones.

The intermediate characteristic is a degenerate triangular disc (collapsed cone) in three dimensions and a pair of lines in two dimensions. For the fast locus we obtain a real cone (or pair of lines) when $\mathbf{u}$ is outside the locus. The slow locus yields two distinct real cones if $\mathbf{u}$ lies inside either of the two cusped regions and a single real cone everywhere else; (a slow cone which has a trace on both cusped regions also contains a triangular "collapsed" disc which joins the two pieces). Summarizing, the sixth order system (neglecting the entropy-induced real streamline characteristic) has three real cones if the velocity lies outside the fast locus or inside the cusped regions (only the former is clearly supersonic). There are only two real cones (and one elliptic imaginary cone) if the velocity falls between the cusped regions and the fast locus. One can distinguish a "fast" from a "slow" Mach cone when P lies within a cusped region by analytic continuation, but it is not clear that this has any significance.

Since there is no substantial general theory, we shall be content to mention a few illustrative examples. First we recall the elementary problem of the linearized supersonic gas flow past a thin airfoil. This is translated as the problem of solving the wave equation in a slit domain (Fig. 8a) with a given boundary condition on the slit (describing the shape of the airfoil). A unique result is obtained only after one imposes an additional regularity condition at infinity upstream. But the mathematical boundary value problem would be equally well-posed if the regularity condition were im-

([1]) In doubtful or singular cases, one should take the envelope of tangent planes rather than tangent lines.

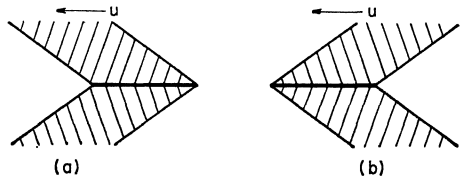


Fig. 8. Airfoil flow.

posed at infinity downstream (Fig. 8*b*), and one can even satisfy the boundary condition on the slit with an arbitrary linear combination of the two solutions indicated in Figs. 8*a* and 8*b*. Of course, one must turn to the time-dependent problem to obtain the correct regularity condition at infinity. The interesting point is that with a certain choice of parameters, there is a linearized *magnetic flow* problem past an airfoil which reduces to the same mathematical problem just described, but the correct choice of the domain of dependence lies *upstream* of the body [17].

Fully hyperbolic problems can be solved using standard means in two dimensions (if linear, using Riemann invariants [16]), and should offer no qualitative surprises in three dimensions [19], except for the unusual orientation and strange shapes of the Mach cones [18]. Although the equations are not unfamiliar, experience from fluid dynamics may be irrelevant. For example, the flow around a bend cannot be solved in terms of shocks or simple waves. The reason is that the boundary condition (at an insulating wall) that the magnetic field is continuous implies that we must solve a static magnetic field problem (elliptic) in the complementary domain and match it to the fluid problem, rather than apply only a local boundary condition [20]. Another feature of the magnetic problem is that, with decreasing magnetic field, the flow does not approach the ordinary fluid flow uniformly [16].

"Mixed" elliptic and hyperbolic problems are another story. The linearized flow past a thin airfoil in such a mixed regime has been successfully carried out [21]. This solution is based on the observation that the variables can be separated into two sets which satisfy elliptic and hyperbolic equations respectively. Since the boundary conditions are applied to certain linear combinations of the elliptic and hyperbolic variables, it becomes necessary to solve an integral equation on the boundary to obtain the solution. We recall that the classical elliptic fluid problem in this slit domain (subsonic flow) is not uniquely determined without specifying a value for the circulation around the body. The magnetic problem is similarly nonunique, but consideration of the singularities leads to a plausible requirement that applies a Kutta condition sometimes at the leading edge and sometimes at the trailing edge, depending on the parameters.

To complete the picture we remark on the possibility that we may encounter transitions of type in different domains ("mixed" or transonic flows in aerodynamic terminology). The combination of mixed equations (partly elliptic, partly hyperbolic in a given domain) together with changes of type from one part of the domain to another seems to offer limitless possibilities for complications. A special case, which is simple enough to yield some information will be discussed in the next section. From the characteris-

tic diagram Fig. 7, we observe that both the transverse and slow Mach cones collapse upon the streamline when $\mathbf{u}$ is parallel to $\mathbf{B}$. If this alignment is satisfied in the entire flow, then certain variables can be integrated out of the problem (just as we can take $\eta = \text{constant}$ in the entire flow), and we are left with a second order system. But from Fig. 7 we easily see that this second order system yields three transitions—elliptic to hyperbolic to elliptic to hyperbolic—as u increases from zero to a large value.

5. Aligned flow

Consider the special case of a flow in which the vectors $\mathbf{u}$ and $\mathbf{B}$ are aligned throughout the flow. From (1.4) we see that this special restriction is significant only in a steady flow. For simplicity, we look first at an incompressible flow and set

$$\mathbf{B} = \sigma \mathbf{u} \tag{5.1}$$

Since div $\mathbf{B} = \text{div } \mathbf{u} = 0$, we conclude that σ is constant on a streamline. By inspection of

$$\varrho(\mathbf{u} \cdot \nabla)\,\mathbf{u} + \nabla(p + B^2/2\mu_0) = \frac{1}{\mu_0}(\mathbf{B} \cdot \nabla)\,\mathbf{B} \tag{5.2}$$

we recognize the analogue variables

$$\left.\begin{aligned} p_* &= p + B^2/2\mu_0 \\ \varrho_* &= \varrho - \sigma^2/\mu_0 \end{aligned}\right\} \tag{5.3}$$

which satisfy the ordinary fluid system

$$\left.\begin{aligned} \varrho_*(\mathbf{u} \cdot \nabla)\,\mathbf{u} + \nabla p_* &= 0 \\ \text{div } \mathbf{u} &= 0. \end{aligned}\right\} \tag{5.4}$$

It is not necessary to refer to an equation of state in an incompressible problem. The pressure can be eliminated from the problem unless it appears explicitly, e.g. in a boundary condition as in the problem of water waves under gravity. The analogy extends to such problems with a free surface in air since the "total" pressure p_* acts exactly as does p an at interface which is a flux surface. Of course, a specifically electromagnetic boundary condition (e.g., that caused by a vacuum magnetic field in the exterior domain) would destroy the analogy.

A slightly more complicated identification of variables yields a similar analogue in a compressible parallel flow [22], [16]. This time $\text{div}(\varrho\mathbf{u}) = 0$ and we set

$$\mathbf{B} = \alpha\varrho\mathbf{u}, \tag{5.5}$$

where α is constant on a stream line. If we insist on

$$\varrho_* \mathbf{u}_* = \varrho\mathbf{u}, \tag{5.6}$$

we immediately detect, from (5.2), the identifications

$$\left.\begin{aligned} \mathbf{u}_* &= \mathbf{u}(1 - \alpha^2\varrho/\mu_0) = \mathbf{u}(1 - 1/M^2), \\ \varrho_* &= \varrho/(1 - \alpha^2\varrho/\mu_0) = \varrho/(1 - 1/M^2), \end{aligned}\right\} \tag{5.7}$$

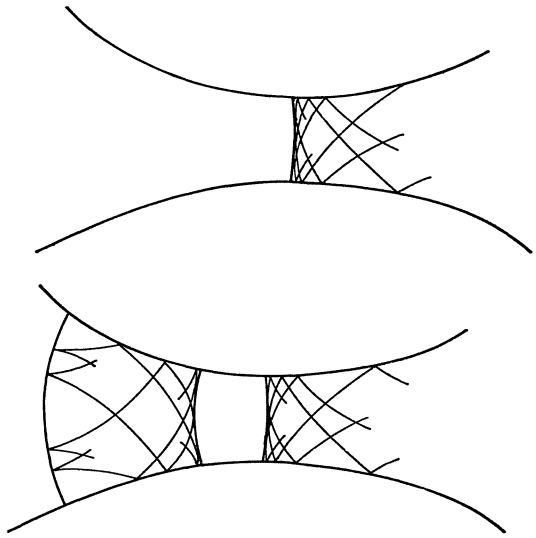


Fig. 9. Nozzle flow.

where M is the Alfvén "Mach" number (a function of ϱ and α)

$$M^2 = u^2/A^2 = \mu_0/\alpha^2\varrho. \tag{5.8}$$

For a compressible flow we *do* require an equation of state. We are forced to accept an equation giving p_* as a function of ϱ_* on each individual streamline, rather than universally. It is obtained by appealing to Bernoulli's law which is valid in the original variables (and also in the new variables with a suitable definition of the enthalpy),

$$h + \tfrac{1}{2}\, u^2 = h_0, \tag{5.9}$$

where h_0 is constant on a streamline. The "equation of state" is

$$p_*(\varrho_*; \eta, \alpha, h_0) = p(\varrho, \eta) + \frac{\alpha^2}{\mu_0}[h_0 - h(\varrho, \eta)]\,\varrho^2 \tag{5.10}$$

after ϱ is eliminated in favor of ϱ_* and α from (5.7). The analogue sound speed is

$$a_*^2 = \partial p_*/\partial \varrho_* = (1 - 1/M^2)^2\,[a^2(1 - 1/M^2) + A^2], \tag{5.11}$$

and the entire analogue thermodynamic structure can be computed [16], but it is evidently unconventional for $M < 1$ (ϱ_* is negative and the analogue flow is in the opposite direction to the actual flow) and even more so for $M^2 < a^2/(a^2 + A^2)$ (the analogue sound speed is imaginary).

Thus we see that the analogue is useful only to a limited extent in studying parallel flow. In particular, it sheds no light on the unusual transitional flows mentioned at the close of Sect. 4. Let us, as an example, consider the flow in a converging-diverging channel. In Fig. 9 we show the qualitative features of a normal transitional flow in a gas and a conjectured magnetic

parallel flow; the characteristics are sketched in the hyperbolic regions. We recall that there are three transitions—elliptic to hyperbolic to elliptic to hyperbolic; these occur at $u^2=a^2$, $u^2=A^2, u^2=a^2A^2(a^2+A^2)$. The second transition, $M^2=1$, is a singularity of the transformation (5.7), and the third transition indicates the onset of the imaginary analogue sound speed. Therefore the normal sonic transition, $u^2=a^2$, is the only one that is not complicated by the introduction of the analogue variables.

Some formal analysis has been done indicating that the flow sketched in Fig. 9 may indeed be legitimate [23], but a mathematically satisfactory answer to this multiple transition question is not yet at hand.

It is interesting to note that the elementary hydraulic approximation to the channel flow in the original variables gives the identical answer as in ordinary gas dynamics; the only facts that are employed in this approximation are the conservation of mass, constant entropy, and Bernoulli's law. In other words, provided that this unusual and difficult flow problem has a solution, the answer is closely approximated by the well-known ordinary gas flow.

A "parallel-transverse" flow in which a two-dimensional aligned flow has a third component of B superposed (in the ignorable direction), can also be transformed into an ordinary fluid problem [16]. But this time the flow blows up near $M=1$ even in the original variables. This indicates that a transition across $M=1$ in the simple aligned flow is quite special.

6. Static equilibrium and force free fields [24]

The problem of static equilibrium is in principle a special case of the previous theories, but they are, in fact, completely unrelated. Fortunately, in the present case there is an exact analogue with a problem in fluid dynamics. But progress in the magnetic problem has carried it beyond the earlier status of the fluid version, so the analogy is, in essence, reversed. The equations are

$$\left.\begin{aligned} &\nabla p=\mathbf{J}\times\mathbf{B},\\ &\mathbf{J}=\frac{1}{\mu_0}\operatorname{curl}\mathbf{B},\\ &\operatorname{div}\mathbf{B}=0. \end{aligned}\right\} \tag{6.1}$$

By inspection we see that the pressure is constant along a magnetic line and also along a current line; or we can say that a pressure surface is covered by magnetic lines and by current lines. In the form

$$\left.\begin{aligned} &\nabla p_*=\frac{1}{\mu_0}(\mathbf{B}\cdot\nabla)\mathbf{B}\\ &p_*=p+B^2/2\mu_0\\ &\operatorname{div}\mathbf{B}=0 \end{aligned}\right\} \tag{6.2}$$

we recognize an exact equivalence with the steady rotational flow of an incompressible fluid(1)

(1) This analogue was first noted by Lundquist [2].

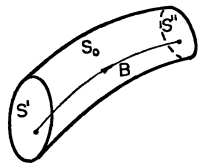


Fig. 10. Tubular equilibrium.

$$\left.\begin{aligned} &\varrho(\mathbf{u}\cdot\nabla)\,\mathbf{u}+\nabla\hat{p}=0,\\ &\operatorname{div}\mathbf{u}=0. \end{aligned}\right\} \tag{6.3}$$

The velocity and magnetic field are identified (within a factor); the "total" pressure p_* is identified with $-\hat{p}$ or, what is equivalent, $-p$ is identified with the Bernoulli constant $\hat{p}+\frac{1}{2}\varrho u^2$.

By combining plausible arguments with proved special cases, we are able to discuss the theory of this system. Consider the tubular domain D of Fig. 10. Roughly speaking, a well-posed boundary value problem will result from the specification of B_n on the boundary of D (say $B_n>0$ on S', $B_n=0$ on S_0, and $B_n<0$ on S''), and also the specification of p on the two ends S' and S''. To be more precise, we suppose that values of B_n and p are given at the end S'. We then take an arbitrary smooth one-one mapping of S' onto S''. The mapping is used to carry the given values of p over to S'', and B_n is multiplied by the Jacobian of the mapping so that $B_n' dS' = B_n'' dS''$. The statement is then that a solution of (6.1) exists which takes the given boundary values of B_n and p and also has the property that the two ends of each magnetic line are related by the given mapping between S' and S''.

The significance of the identification of the ends of the magnetic lines can be made clear by two examples. If the lines $p=$constant cover S' simply (Fig. 11*a*), then the specification of B_n independently on S' and S'' subject to $\int B_n dS' + \int B_n dS'' = 0$) and the specification of compatible p values at both ends (i.e.,

$$\int_{p<p_0} B_n\,dS' + \int_{p<p_0} B_n\,dS'' = 0$$

for all p_0) uniquely determines a mapping of S' onto S''. On the other hand, if the lines $p=$constant are closed curves (Fig. 11*b*), then specification of compatible boundary values of B_n and p at both ends does not fix the mapping. In addition to these boundary values, we also specify a given "twist" from S' to S'' of each constant p curve.

The necessity for this twist becomes clear when we note that a vector field $\mathbf{B}$ taken from a solution of (6.1) is, in a certain generalized sense, a

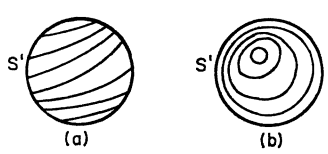


Fig. 11. Pressure boundary condition.

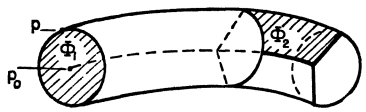


Fig. 12. Toroidal equilibrium.

harmonic vector field within a pressure surface [24]. In particular, such a surface harmonic is uniquely determined by boundary values when the pressure surface is simple, Fig. 11*a*, but it requires the specification of certain *periods* (e.g., circulation) when the pressure surface is not simple, Fig. 11*b*.

Making use of this concept, we are led to a formulation in a toroidal domain. In this case there are no boundaries (except the trivial one on which $B_n=0$) and we can specify *only* periods on each pressure surface. Specifically, a solution of the system (6.1) with nested toroidal pressure surfaces can be expected to be determined by the specification of two functions $\Phi_1(p),\Phi_2(p),p_0<p<p_1$, which give the total flux intercepted in the two independent directions between the axis p_0 (a closed curve) and a given surface p (Fig. 12). An alternative formulation is to fix the two fluxes $\Phi_1(V)$, $\Phi_2(V)$ as functions of the volume V which parametrizes the nested toruses (p will be determined only within an added constant). It should be remembered that the family of nested toruses must be found as well as the associated $\mathbf{B}$ and p; only the outer boundary is known beforehand.

The special case of (6.1) with $p=0$,

$$\left.\begin{aligned} \operatorname{curl} \mathbf{B}\times\mathbf{B}&=0,\\ \operatorname{div} \mathbf{B}&=0, \end{aligned}\right\} \tag{6.4}$$

is called a *Force-Free Field* (in the fluid version, the Bernoulli constant is constant over the entire flow and we have a Beltrami field). This is a significant concept in the theory of ionized gases because it is possible for a gas to be very rarefied ($p\sim 0$) and still carry an appreciable current. In the tubular domain, Fig. 10, the statement of a well-posed problem is the same as before, omitting the specification of p. A mapping of S' onto S'' is given and also the value of B_n; a solution will take the given value of B_n and will satisfy the end coordination. In a torus, p serves only as a parameter to identify the nested toruses. There is no difference between the pressure and volume parametrization, and the parametric representation $\Phi_1(p),\Phi_2(p)$, $p_0<p<p_1$, is equivalent to specifying Φ_1 as a function of Φ_2 over a specified range $0<\Phi_2<\Phi^*$.

It is illuminating to compare a general equilibrium (6.1), a Force-Free Field (6.4), and a harmonic vector (curl $\mathbf{B}=0$, div $\mathbf{B}=0$) in the domain between two concentric toruses. For the system (6.1) we specify $\Phi_1(p)$ and $\Phi_2(p)$; for the system (6.4) we specify only $\Phi_1(\Phi_2)$; for the harmonic field we specify two constants, Φ_1 and Φ_2.

This conjectured existence theory quoted above is based on a study of the characteristics of the system (6.1), on a number of proved special cases, and on deductions from a variational formulation of the problem. The system has both real and imaginary characteristics. But in contrast with the mixed elliptic and hyperbolic problems which appear in steady flow problems, the

'eal cones are all degenerate and fall on the streamlines. This suggests the ıpecification of B_n as a boundary condition and p as a separate "initial :ondition" at each end but the conditions involving periods on nonsimple ›ressure surfaces can only be discovered after the interpretation of B as a ıurface harmonic. The solvable special cases are two-dimensional and ıxially symmetric problems (and include the toroidal case by imposing ›eriodicity on an axially symmetric problem). The variational formulation 'or the tubular domain, Fig. 10, is to make the variational function

$$\int (B^2/2\mu_0 - p)\, dV \tag{6.5}$$

ıtationary subject to an admissible class of functions (B,p) satisfying

$$\left.\begin{aligned} \operatorname{div} \mathbf{B} &= 0 \\ \mathbf{B}\cdot\nabla p &= 0 \end{aligned}\right\} \tag{6.6}$$

;ogether with the boundary conditions on B_n and p and the mapping 'estriction at the ends of the magnetic lines. Under these conditions an nterior variation yields the Euler–Lagrange equation $\nabla p = \mathbf{J}\times\mathbf{B}$, and there s no contribution from the boundary variation.

In the toroidal problem, the same variational function (6.5) is used and ın admissible (B,p) satisfies (6.6) as well as the given flux requirements $\mathfrak{P}_1(p),\Phi_2(p)$. In other words, an admissible pair (B,p) is obtained from a set ›f nested toruses by assigning values of p and a tangential magnetic field :ompatible with the given flux. In the case of given $\Phi_1(V),\Phi_2(V)$, we use

$$F = \int B^2/2\mu_0\, dV \tag{6.7}$$

ıs the variational function. Given a set of concentric toruses, the parameter V is known and we choose a tangential field B compatible with the given 'luxes. The pressure does not enter this variational formulation and takes :he form of a Lagrangian multiplier in (6.5) if we wish to specify $\Phi_1(p)$, $\mathfrak{P}_2(p)$ instead of $\Phi_1(\mathrm{U}),\Phi_2(\mathrm{U})$.

As a final illustration, let us consider the classical Dirichlet principle for :he tubular domain, Fig. 10. A harmonic vector is uniquely determined by :he given boundary condition on B_n. Variationally, we minimize F in (6.7) ıubject to the admissibility condition

$$\left.\begin{aligned} \operatorname{div} \mathbf{B} &= 0 \\ B_n &= \text{given} \end{aligned}\right\} \tag{6.8}$$

;o obtain the variational condition

$$\operatorname{curl} \mathbf{B} = 0. \tag{6.9}$$

A simple way of incorporating the condition div $\mathbf{B}=0$ is to introduce the 'ector potential

$$\mathbf{B} = \operatorname{curl} \mathbf{A}, \tag{6.10}$$

after which (6.9) is obtained by the usual integration by parts. But, if we parametrize div $\mathbf{B}=0$ by

$$\mathbf{B}=\nabla\alpha\times\nabla\beta \tag{6.11}$$

(the existence of such α and β is both necessary and sufficient for div $\mathbf{B}=0$), an integration by parts in the variation of F yields the interior variation

$$\operatorname{curl}\mathbf{B}\times\mathbf{B}=0 \tag{6.12}$$

instead of curl $\mathbf{B}=0$. Since the variational function is the same as in Dirichlet's principle and the variational condition (6.12) is wider, the admissibility class must be (unwittingly) more restricted. Indeed, performing the boundary variation we discover that the boundary condition $B_n=$given is not sufficient to make F stationary. Fixing α and β at both ends of the tube does make F stationary, and this is exactly the mapping condition on the ends of the magnetic lines (both α and β are constant along a magnetic line).

The resolution of the surprising distinction between the two parametrizations of div $\mathbf{B}=0$ is that, although they are equivalent in the large, a local variation of α and β keeps intact the identification of magnetic lines across the varied domain. In the general equilibrium problem, (6.5), the parametrization

$$B=\nabla p\times\nabla\omega \tag{6.13}$$

(where ω may be multivalued) accomplishes the same purpose.

7. Free boundary equilibrium and stability [25]

We now turn to a problem which was one of the first in magneto-fluid dynamics to be studied in depth. The reason this could be done is that the electromagnetic field is almost completely separated from the fluid dynamics. Specifically, we consider an interface which separates a pure electromagnetic field in vacuo from a conducting fluid within which there is no electromagnetic field. By the perfect conductivity, this separation persists in time. The fluid provides boundary conditions for the electromagnetic field, viz.

$$\left.\begin{aligned} B_n&=0\\ (\mathbf{E}+\mathbf{u}\times\mathbf{B})_t&=0\end{aligned}\right\} \tag{7.1}$$

and the field suplies a boundary condition for the fluid motion

$$p=B^2/2\mu_0. \tag{7.2}$$

To provide a suitable formulation it is necessary to use a Galilean invariant form of the electromagnetic equations [26]. The result is a theory which is basically elliptic (in particular, $\mathbf{B}$ is determined at any instant by the instantaneous geometrical configuration plus the specification of certain, possibly time-varying, periods such as fluxes or currents). Heuristically, we conclude that the driving pressure (7.2) on the fluid is determined by the instantaneous position of the interface; thus the fluid motion is determined.

The simplest of these problems is the static equilibrium of a fluid at rest

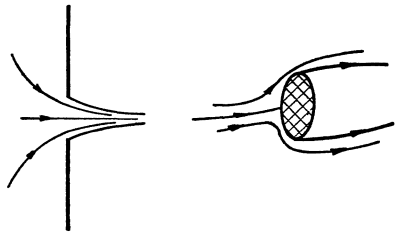


Fig. 13. Free boundary flows.

(p = constant) balanced by a magnetic field. This is, of course, a special case of the static equilibria studied in the previous section. It is therefore an analogue of a fluid flow problem, viz., the classical free boundary flow. In this problem we have a potential flow, div $\mathbf{u}=0$, curl $\mathbf{u}=0$, separated at a stream surface from either stagnant water or a cavity. The additional boundary condition, $|u|$ = constant at the interface (obtained from Bernoulli's law) serves to determine the shape of the interface. In the magnetic version, the vacuum field curl $\mathbf{B}=0$, div $\mathbf{B}=0$ is also provided with an extra boundary condition, $|\mathbf{B}|$ = constant at the interface.

Although the two problems are mathematically identical, the solutions which are of interest in the two cases are not the same. For example, the fluid flow from an orifice and the cavitational flow around an obstacle (Fig. 13) are of very little interest magnetically. On the other hand, important magnetic configurations such as the *pinch* and the *cusped geometry* (Fig. 14) are unimportant fluid dynamically. One important reason for this distinction is the question of stability. The stability problem involves time-dependent equations which are *not* analogous in the magnetic and fluid versions. A very complete analysis of the magnetic free boundary stability problem has been done. The end result is a simple criterion which is necessary and sufficient for stability against finite amplitude displacements in terms of easily verified properties of the equilibrium configuration. The great generality of this theory results from the existence of a variational formulation. The appropriate variational function is

$$F=\int_{D_v}(B^2/2\mu_0)\,dV+\int_{D_f}\varrho e\,dV-\sum_1^m I_r\Phi_r, \tag{7.3}$$

where the first integral (over the vacuum domain) is the magnetic energy, the second integral (over the fluid domain) is the fluid energy, and the summation is carried over those magnetic circuits $1\ldots m$ for which the currents I_r are fixed constraints; in the remaining magnetic circuits, $m+1,\ldots,n$, the

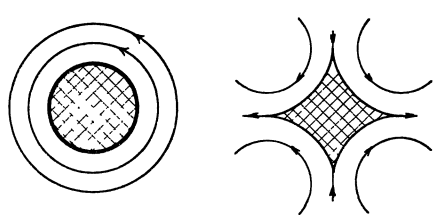


Fig. 14. Magnetic free boundary solutions.

(a) (b)

Fig. 15. Stable equilibria.

fluxes Φ_r are fixed constraints. An equilibrium configuration is characterized by stationary F;(1) a stable configuration is characterized by a minimum value of F.

Specifically, the theorem [25](2) is that a given configuration is stable if there exists a domain D_1 which encloses the fluid domain D_0 and within which the magnitude B of the magnetic field is larger than the constant value B_0 taken on the interface, $B > B_0$ in $D_1 - D_0$. It is unstable if there is a region adjacent to D_0 where $B < B_0$. This criterion is valid for any geometrical and topological fluid and confining coil configuration; for any position of external fixed conductors, touching the fluid or not; for constant flux or constant current constraints (or any combination) on the confining fields; for any fluid properties, incompressible or compressible with any equation of state, or even subject to the Boltzmann equation. The essential mathematical content of this problem is the domain variation of harmonic vectors in regions of higher topological structure.

There is a simple relation between the above criterion for stability in the large and the more elementary criterion for positivity of the second variation of the appropriate variational function. The latter condition is simply that $\partial B^2/\partial n$ be positive on the entire interface for stability, and negative anywhere for instability. This criterion has a simple geometrical interpretation. Since

$$\operatorname{curl} \mathbf{B} \times \mathbf{B} = (\mathbf{B} \cdot \nabla)\mathbf{B} - \nabla(\tfrac{1}{2} B^2) = 0 \tag{7.4}$$

for a harmonic vector, the sign of $\partial B^2/\partial n$ is associated with the sign of the curvature of a magnetic line on the interface (the magnetic lines are easily seen to be geodesics, so this is the curvature of the surface in the given direction). From this remark one can easily prove the theorem that no finite fluid domain with a smooth boundary can be stable (at a point of tangency with a support plane, the curvature must be in the unstable direction). This stability analysis led to the discovery of the stable cusped configuration of Fig. 14.

There are many stable three-dimensional configurations [29] of which we indicate only two in Fig. 15. These configurations are stable once they have been shown to exist. Two-dimensional equilibria are easily found by conformal mapping. The periodic structure, Fig. 15*b*, can be shown to exist by perturbation from the known two-dimensional solution, but the singularity on the axis of Fig. 15*a* has so far eluded an existence proof (although the figure has been empirically computed, numerically).

It is instructive to compare several common definitions of stability:

(1) boundedness of solutions of the nonlinear partial differential equations,
(2) verification of a minimum in the large for F, eq. (7.3),

(1) This is a modification of a variational principle used by Friedrichs to show existence of an axially symmetric fluid free boundary [27].

(2) For a more complete account see [28].

(3) determination of the sign of the second variation of F,
(4) determination of the characteristic frequencies of a linearized solution of the partial differential equations, assumed to be exponential in time.

The approach (1) has not been attempted. The relation between (2) and (3) is the simple one stated above. There is also a simple connection between (3) and (4); there is a variational formulation for the charcteristic frequencies which uses the second variation of F, just as in classical vibration problems.

In this last formulation, the free boundary stability problem has been generalized (for the linearized problem) to many cases of general equilibria, $\nabla p = \mathbf{J} \times \mathbf{B}$ [30]. This problem is more complicated than the free boundary because it can be shown that there is no local property of the equilibrium (as there is in the free boundary case) that can characterize stability [25]. In particular, there are no results concerning finite amplitude stability.

Historical note

Many of the early results in this subject (prior to 1958) were first presented in restricted conferences concerned with Project Sherwood controlled thermonuclear research)(1). Most of the fluid analogues (transverse flow, Sect. 3, incompressible parallel flow, Sect. 5, general static(2) and free boundary equilibria, Sects. 6 and 7), also the variational formulation of the many static equilibria, Sect. 6 (including a tentative formulation for the toroidal geometry) as well as the variational formulation of the linear and nonlinear free boundary stability problem, Sect. 7, were presented by the author in lectures at New York University,(3) Livermore,(4) and Princeton(5) in 1954. The Galilean invariant electromagnetic theory [26] was developed along two different lines by K. O. Friedrichs and the author in 1954.(6) The characteristic theory, Sect. 2, was presented by K. O. Friedrichs in lectures at New York University, Los Alamos,(7) and Princeton(8) in 1954. The theorem relating the linear stability of a free boundary configuration to the curvature of the boundary as well as the theorem of nonexistence of smooth stable equilibria, Sect. 7, were announced and proved by the author in 1955.(9) An independent heuristic stability argument involving particle orbits leading to a roughly similar criterion was given by Rosenbluth and Longmire [31]. The variational free boundary stability analysis was subsequently extended by E. Frieman and others to general equilibria (without an

(1) An excellent general account is given in A. S. Bishop, *Project Sherwood*, Addison-Wesley Publishing Co., 1958.
(2) This analogue was noted earlier by Lundquist [2].
(3) Mimeographed lecture notes on magnetohydrodynamics (I. *General Equations*, II, *Transients*, III. *Ohm's Law*, IV. *Potential Flows and Similarities with Fluid Dynamics*), New York University, August 16, 1954.
(4) *Note on Magnetohydrostatics* (mimeographed), Livermore, August 30, 1954.
(5) U.S. Atomic Energy Commission Report WASH-184, p. 144 (issued January 1955).
(6) WASH-184, pp. 144 and 148.
(7) *Nonlinear Wave Motion in Magnetohydrodynamics*, Los Alamos Report LAMS-2015 (written Sept. 1954, issued 1957).
(8) WASH-184, p. 148.
(9) U.S. Atomic Energy Commission Report WASH-289, p. 115 (issued June 1955).

interface) for the vibrational form of the linear problem [30](1) (also independently in Germany [32] and in the U.S.S.R. [33]). The existence of the two-dimensional cusped equilibrium was shown by Friedrichs.(2) The non-existence of a simple local stability criterion in any but a free boundary equilibrium was shown by H. Rubin in 1955.(3) The free boundary stability theorem in the large was established by the author and A. A. Blank in 1955.(4) The toroidal equilibrium theory was further developed by the author and H. Rubin in 1955 and 1956 [24], and a somewhat similar simultaneous development was carried out by Kruskal and Kulsrud [34].

Acknowledgement

The work presented in the paper was supported by the Courant Institute of Mathematical Sciences, New York University, under Contract AT(30–1) –1480 with the U.S. Atomic Energy Commission.

REFERENCES

[1]. GÅRDING, L., *Acta Math.*, 85 (1951), 1.
[2]. LUNDQUIST, S., *Ark. Fysik*, 5 (1952).
[3]. REICHEL, P., New York Univ. Report NYO-7697, April 1958.
[4]. ZUMINO, B., *Phys. Rev.*, 108 (1957), 1116.
[5]. BLANK, A. A. & GRAD, H., New York Univ. Report NYO-6486-VI, July 1958.
[6]. FRIEDRICHS, K. O. & KRANZER, H., New York Univ. Report NYO-6486-VIII, July 1958.
[7]. LYNN, Y. M., *Phys. Fluids*, 5 (1962), 627.
[8]. BAZER, J. & FLEISCHMAN, O., *Phys. Fluids*, 2 (1959), 366.
[9]. GRAD, H., Propagation of magnetohydrodynamic waves without radial attenuation. In: *The Magnetodynamics of Conducting Fluids* (ed. D. BERSHADER). Stanford University Press, 1959.
[10]. LUDWIG, D., New York Univ. Report NYO-9351, January 1961.
[11]. ALFVÉN, H., *Cosmical Electrodynamics*. Oxford, 1953.
[12]. WEITZNER, H., *Phys. Fluids*, 4 (1961), 1238.
[13]. FRIEDLANDER, F. G., *Proc. Cambridge Philos. Soc.*, 55 (1959), 341.
[14]. LIGHTHILL, M. J., *Phil. Trans. Roy. Soc.* (London), A252 (1960), 397.
[15]. GARDNER, C. S., New York Univ. Report NYO-9354, September 1960.
[16]. GRAD, H., *Rev. Mod. Phys.*, 32 (1960), 830.
[17]. SEARS, W. R., *Rev. Mod. Phys.*, 32 (1960), 701.
[18]. CUMBERBATCH, E., *J. Aerospace Sci.* 29 (1962), 1476.
[19]. SARASON, L., *Comm. Pure Appl. Math.*, 15, 4 (1962).

---

(1) WASH-289, p. 343. This formulation, for vibrational stability, was originally introduced by Lundquist [2].

(2) WASH-289, p. 115. A somewhat similar configuration had been described earlier in the fluid version, as a collision of jets but was dismissed (properly) as nonphysical.

(3) U.S. Atomic Energy Commission Report TID-7503, p. 247 (issued February 1956).

(4) TID-7503, p. 238.

[20]. CHU, C. K. & LYNN, Y. M., Steady magnetohydrodynamic flow past a nonconducting wedge. *J. Am. Rocket Soc.* (To appear.)

[21]. CUMBERBATCH, E., SARASON, L. & WEITZNER, H., *J. Aerospace Sci.* 29, 2 (1962).

[22]. IMAI, I., *Rev. Mod. Phys.*, 32 (1960), 992.

[23]. CHU, C. K., *Phys. Fluids, 5* (1962), 550.

[24]. GRAD, H. & RUBIN, H., *Proc. of 2nd International Conference on the Peaceful Uses of Atomic Energy*, Geneva, 1958 Vol. 31, p. 190.

[25]. BERKOWITZ, J., GRAD, H & RUBIN, H., *Proc. of 2nd International Conference on Peaceful Uses of Atomic Energy*, Geneva, 1958, Vol. 31, p. 177.

[26]. BLANK, A. A., FRIEDRICHS, K. O. & GRAD, H., New York Univ. Rep. NYO-6486-V, November 1957.

[27]. FRIEDRICHS, K. O., *Math. Ann.*, *109* (1933), 60.

[28]. BLANK, A. A., Stability of magneto-fluid free boundaries against finite perturbation. (To appear in *Proc. Symposium on Fundamental Topics in Fluid Mech.*, Michigan State University, October 1962.)

[29]. BERKOWITZ, J., *et al.*, *Proc. of 2nd International Conference on the Peaceful Uses of Atomic Energy*, Geneva, 1958, Vol. 31, p. 171.

[30]. BERNSTEIN, I. B., *et al.*, *Proc. Roy. Soc.*, A244 (1958), 17.

[31]. ROSENBLUTH, M. N. & LONGMIRE, C. L., *Ann. Phys.*, 1 (1957), 120.

[32]. HAIN, K., LÜST, R. & SCHLÜTER, A., *Z. Naturforschg.*, 12a (1957), 833.

[33]. BRAGINSKII, S. I. & KADOMTSEV, B. B., Stabilization of a plasma by the use of guard conductors. In: *Plasma Physics and the Problem of Controlled Thermonuclear Reactions*, Vol. III, p. 356. Pergamon Press, 1959.

[34]. KRUSKAL, M. D. & KULSRUD, R. M., *Proc. of 2nd International Conference on the Peaceful Uses of Atomic Energy*, Geneva, 1958, Vol. 31, p. 213.

NEW RESULTS ON THE STABILITY OF PERIODIC MOTIONS

By JÜRGEN MOSER

This lecture has been presented in a similar form at the Symposium on Nonlinear Problems at Madison, April 1962 and will be published in the Proceedings of that conference. To avoid duplication we only summarize the main theorems of this paper:

The first theorem refers to the existence of invariant curves of a mapping defined in an annulus $1 \leqslant r \leqslant 2, \theta (\mathrm{mod}\, 2\pi)$ where r, θ are polar coordinates. The mapping considered has the form

$$M: \qquad \theta_1 = \theta + \alpha(r) + f(\theta, r),$$
$$r_1 = r + g(\theta, r),$$

where f, g are small, in a sense to be specified below. The function $\alpha(r)$ is assumed to be monotone so that $|\alpha(2) - \alpha(1)| = \Delta$ is positive. For any function $h(r, \theta)$ we define $|h|_l$ as the maximum of all derivatives of order $\leqslant l$ taken over the total annulus. We make the following further assumptions:
(1) There exists a constant $c > o$ such that

$$c^{-1}\Delta \leqslant |\alpha'|, \quad |\alpha'|_l \leqslant c\Delta \leqslant c^2$$

(2) Every closed curve $C: r = \varphi(\theta)$ with $|\varphi'| \leqslant 1$ intersects its image curves MC.

THEOREM 1. *For $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon, c)$ such that for*

$$\frac{|f|_l + |g|_l}{\Delta} < \delta \quad \text{and} \quad l \geqslant 333$$

the mapping M possesses an invariant curve

$$r = \psi(\theta) \textit{ with } |\psi'| < \varepsilon.$$

The proof of this theorem appeared in *Nachr. Akad. Wiss. Göttingen*, II, math.-phys. Kl. 1962, No. 1. There also the relevant literature, especially the papers of Kolmogorov and Arnold, is quoted.

Theorem 1 can be used for stability discussions of periodic solutions Hamiltonian systems of two degrees of freedom. This problem can be reduced to the study of an area preserving mapping

$$(*) \qquad z_1 = f(z, \bar{z}) = \lambda z + F(z, \bar{z})$$

with
$$|\lambda| = 1, \quad \frac{(\partial f, \bar{f})}{\partial(z, \bar{z})} \equiv 1 \quad (z = x + iy)$$

near a fixed point $z = o$. The function F is assumed to vanish quadratically:

$$F = O_{l'}(|z|^2);$$

here we use the notation $G = O_l(|z|^s)$, if there exists a constant $c > 0$ such that for $\varrho + \sigma \leqslant l$ one has

$$\frac{|z|^{\varrho+\sigma}}{\varrho!\,\sigma!}\left|\frac{\partial^{\varrho+\sigma} G}{\partial z^\varrho \partial \bar{z}^\sigma}\right| \leqslant c\,|z|^s \quad \text{in} \quad 0 < |z| < c^{-1}.$$

As a simple generalization of a theorem of G. D. Birkhoff (1920) we obtain

THEOREM 2. *If $\lambda, \lambda^2, \ldots, \lambda^q \neq 1$ there exists a real analytic coordinate transformation of (x,y) into (u,v) with non-vanishing Jacobian such that the mapping M takes the form*

$$(**) \qquad w_1 = w e^{i(\alpha+\beta|w|^s)} + O_l(|w|^q) \quad (w = u + iv),$$

where s is an even integer with $0 < s < q-1$ and $\beta = 0, \pm 1$ (independently of the choice of $q > s+1$).

Here it is assumed that in the original mapping

$$F = O_{l'}(|z|^2) \quad \text{with} \quad l' \geqslant l + q - 2.$$

The main result is contained in

THEOREM 3. *If $\beta \neq 0$ then the fixed point $w = 0$ is stable under the mapping (**), provided $l \geqslant 333$.*(1)

In general one can choose $q = 4$, $s = 2$ in which case the stability conditions take the form

$$\lambda^3 \neq 1, \quad \lambda^4 \neq 1, \quad \beta \neq 0.$$

In conclusion several applications of Theorem 1 were mentioned. In connection with the plane restricted 3 body problem it is an old problem to describe initial conditions which never lead to collisions. If the ratio of the non zero masses, say μ, is small such a condition can be given. More precisely, it can be shown that

$$\operatorname{osc} e(t) \leqslant \mu^{3/8} \quad \text{for} \quad 0 < \mu < \mu_0,$$

where $e(t)$ is the eccentricity of the osculating ellipse for the motion of the zero mass.(2)

Another example is the motion of a charged particle in a magnetic dipole field where the mass of the particle is small. Such a particle performs rapid oscillations about the magnetic field lines. C. S. Gardner was able to find bounds for the solutions of this problem which are valid for all times.

The result of Theorem 1 can be generalized to canonical mappings for several—say $2n$—dimensions. One obtains a set of n dimensional invariant tori. The proof of this result does not lead to new difficulties. However, such a result is not sufficient for stability investigations since one needs the

(1) A similar result was announced by V. I. Arnold, *Dokl. Akad. Nauk*, 137 (1961), 255–257.

(2) The author learned during the discussion that Professor Kolmogorov had found this result and presented it in Moscow and Paris in unpublished form.

existence of $(2n-1)$ dimensional invariant surface and for $n>1$ an n-dimensional is not a boundary. The nonexistence of invariant surface of $(2n-1)$ dimensions for Hamiltonian systems of n degrees of freedom had already been discussed by E. Fermi ("Beweis, dass ein mechanisches Normalsystem im allgemeinen quasi-ergodisch ist". *Phys. Z.* 24 (1923), 261–265) in a nonrigorous fashion. Therefore this approach does not permit the construction of invariant open sets in the case of several degrees of freedom. It is possible, however, to establish the existence of a measurable invariant set of positive measure. A remarkable result in this direction, applied to the many body problem has been presented by V. I. Arnold at this congress.

SOME RESULTS ON THE STRUCTURE OF RELATIVISTIC QUANTUM FIELD THEORY

By A. S. WIGHTMAN

1. Heuristic considerations [1]

The problem of constructing relativistic quantum field theories is about thirty-five years old. In a typical case it may be formulated as follows. Given the classical field equation

$$(\Box + m^2)\, u(x) = g[u(x)]^3,$$

where

$$\Box = \left(\frac{\partial}{\partial x^0}\right)^2 - \sum_{j=1}^{3}\left(\frac{\partial}{\partial x^j}\right)^2 \tag{1}$$

and g is a real constant, for a real valued function, u, defined for all space time points, x, find a corresponding quantum field theory in which u is replaced by a hermitean operator valued function, A, satisfying (1) as an operator equation in a Hilbert space $\mathcal{H}$. Two further restrictions are made on a field operator, A, in quantum field theory. First, it must have the property of *local commutativity*

$$[A(x), A(y)] = 0 \tag{2}$$

for space-like separated x and y (i.e. x and y satisfying

$$(x-y)^2 = (x^0 - y^0)^2 - \sum_{j=1}^{3} (x^j - y^j)^2 < 0).$$

Second, there must exist a continuous unitary representation of the inhomogeneous Lorentz group $\{a,\Lambda\} \to \mathcal{U}(a,\Lambda)$ such that

$$A(\Lambda x + a) = \mathcal{U}(a,\Lambda) A(x) \mathcal{U}(a,\Lambda)^{-1}. \tag{3}$$

Here a is a space-time translation and Λ a homogeneous Lorentz transformation. Because $\mathcal{U}(a,1)$ is necessarily of the form $\exp i P^\mu a_\mu$ and the interpretation of the four-vector operator, P^μ, is as the observable of energy and momentum one requires further that the spectrum of P^μ lie in the future light cone $p^2 \geqslant 0$, $p^0 \geqslant 0$.

It was early realized that A had to be understood as a singular function so that only the smeared field quantities

$$A(\varphi) = \int dx\, \varphi(x)\, A(x) \tag{4}$$

could be given literal meaning as field operators. (Here is any one of a class of smooth functions whose precise properties will be discussed later.) This was already evident in the simplest case $g = 0$ where it is easy to construct a solution, the so-called *free field* which satisfies in addition to

$$(\Box + m^2)\, A(x) = 0, \tag{5}$$

the commutation relation

$$[A(x), A(y)] = \frac{1}{i}\Delta(x-y) \tag{6}$$

which holds for all x and y. Equation (6) is consistent with (2) because the distribution Δ, which is defined by

$$\Delta(x) = \frac{i}{2(2\pi)^3}\int e^{-ik\cdot x}\operatorname{sgn} k_0\, \delta(k^2 - m^2)\, dk \tag{7}$$

vanishes for space-like x. With a suitable specification of the domains of the operators $A(\varphi)$, it turns out that equations (5) and (6) have an irreducible solution unique up to unitary equivalence.[2]

The fact that $A(x)$ must in general be a distribution rather than a function gives rise to the basic problem of giving a meaning to $[A(x)]^3$ in equation (1). As a first step in understanding this problem, one can look at what meaning can be given to $[A(x)]^n$ when $A(x)$ is a free field. This simpler problem was solved in the late twenties by the introduction of what in a modern notation is, in the simple case $n=2$ for example,

$$:A(x)^2: = \lim_{x_1,\, x_2 \to x}\left[A(x_1)A(x_2) - \frac{1}{i}\Delta^{(+)}(x_1 - x_2)\right], \tag{8}$$

where

$$\Delta^{(+)}(x) = \frac{i}{2(2\pi)^3}\int e^{-ik\cdot x}\tfrac{1}{2}(1 + \operatorname{sgn} k_0)\, \delta(k^2 - m^2)\, dk.$$

An analogous definition of $:A(x)^n:$ turns out to yield a mathematical object quite as impeccable as the free field itself. It is suggested by this example that in place of $[u(x)]^3$ in (1) one should have some sort of finite part say $PF([A(x)]^3)$. The problem is how to define it, especially in view of the fact that exploratory investigations seem to show that the : : which works for the powers of the free field is not adequate in general.

After preliminary skirmishes in the 1930's, important progress in this problem was achieved in the late 1940's. It can be described as follows. Convert (1) into the integral equation

$$A(x) = A^{(0)}(x) + \int \Delta_R(x-y)\, g[A(y)]^3\, dy, \tag{9}$$

where $A^{(0)}(x)$ is a free field and $\Delta_R(x)$ is the retarded elementary solution of equation (5):

$$\left.\begin{aligned} (\Box + m^2)\,\Delta_R(x) &= \delta(x), \\ \Delta_R(x) &= 0, \quad (x^0 < 0). \end{aligned}\right\} \tag{10}$$

Then solve (9) by iteration starting with $A^{(0)}(x)$ as a first approximation. It turns out that all terms after the first are infinite. However, the deep discovery was made that a formal separation into finite and infinite parts was possible and that the infinite parts in question could all be ascribed to

(infinite!) changes in m, g, and the scale of the field. This means that if PF is defined

$$gPF([A(x)]^3) = g:A(x)^3: + a\,A(x) + b(\Box + m^2)\,A(x) + gc:A(x)^3:, \tag{11}$$

where a, b, and c are formal power series in g, the coefficients in the power series can be chosen so that the equation

$$A(x) = A^{(0)}(x) + g\int \Delta_R(x-y)\,PF([A(y)]^3)\,dy \tag{12}$$

has an iterative solution free of divergences; this is customarily referred to as the *renormalized perturbation series.*

While this was decisive progress, it can scarcely be regarded as a complete solution of the problem in view of the fact that it could be shown that the solution was a divergent series in g. Clearly, one needed techniques more powerful than a perturbation expansion.

Mathematically, what was in order was a three step program: (a) Isolate a class of $A(x)$ which is sufficiently rich to be interesting and investigate its structure; (b) Show how $PF([A(x)]^3)$ can be defined for a suitable subclass of $A(x)$; (c) Show that (12) can be satisfied within that subclass. From the mathematical point of view, it can be said that the last few years of quantum field theory have been devoted to (a). The notion of field has been axiomatized and quite a number of general theorems have been proved. In fact, it has turned out that so much can be proved directly from the assumed general properties of a field that such specific problems as (b) and (c) have been left by the wayside. The fact is, however, that we are only beginning to get that insight into the consequences of the axioms which would make possible a direct attack on (b) and (c).

The results described below are of two kinds, general theorems valid for every field theory satisfying the axioms and detailed investigations of special cases studied for the light they may throw on the general theory. Results of the latter kind have been so far been rather modest; there is a conspicuous scarcity of non-trivial examples of theories of fields.

2. Axioms for a hermitean scalar field [3]

Let $\mathcal{H}$ be a separable Hilbert space and $\{a,\Lambda\}\to\mathcal{U}(a,\Lambda)$ a continuous unitary representation of the inhomogeneous Lorentz group, whose energy momentum spectrum lies in or on the future light cone. Suppose there is a unique vector Ψ_0. (the vacuum state) satisfying $\mathcal{U}(a,\Lambda)\Psi_0 = \Psi_0$. Let $\mathcal{D}$ be the set of all complex valued infinitely differentiable functions of compact support. Then a hermitean scalar field is a linear function from $\mathcal{D}$ to the linear operators of $\mathcal{H}$ such that

I. (*Properties of the domains of the field operators*)

The operators $A(\varphi)$ and $A(\varphi)^*$, $\varphi\in\mathcal{D}$ possess a common dense domain, D, which is linear and has the properties

$$\left.\begin{array}{l} A(\varphi)D\subset D, \quad A(\varphi)^*D\subset D, \quad \Psi_0\in D \\ U(a,\Lambda)D\subset D. \end{array}\right\} \tag{13}$$

For each pair $\Phi,\Psi \in D$, $(\Phi, A(\varphi)\Psi)$ is a distribution in φ

On D
$$A(\varphi)^* = A(\bar{\varphi}), \tag{14}$$

where $\bar{\varphi}$ is the complex conjugate of φ.

II. (*Transformation law of the field*)

$$\mathcal{U}(a,\Lambda)\, A(\varphi)\mathcal{U}(a,\Lambda)^{-1} = A(\{a,\Lambda\}\varphi) \tag{15}$$

on D. Here
$$(\{a,\Lambda\}\varphi)\,(x) = \varphi(\Lambda^{-1}\,(x-a)).$$

III. (*Local commutativity*)

$$[A(\varphi), A(\psi)] = 0 = [A(\varphi), A(\psi)^*] \tag{16}$$

on D for all pairs $\varphi,\psi \in \mathcal{D}$ and such that $\varphi(x)\,\psi(y) = 0$ for all x and y such that $(x-y)^2 \geqslant 0$.

These axioms have been given for test functions φ in $\mathcal{D}$, the space of infinitely differentiable functions of compact support. They could equally well be given for $\mathcal{S}$, the space of infinitely differentiable functions which, together with their derivatives vanish faster than any power of the distance at infinity. Actually from the physical point of view the observability of fields in bounded regions of space-time supports the weaker assumption ($\mathcal{D}$), but for applications to elementary particle physics it is very natural and convenient to have the stronger assumption ($\mathcal{S}$).

3. Vacuum expectation values as a characterization of a theory of fields [4]

One of the immediate consequences of I is that the quantities

$$(\Psi_0, A(\varphi_i) \ldots A(\varphi_n)\Psi_0) \tag{17}$$

exist and are multilinear functionals of $\varphi_1, \ldots, \varphi_n$ separately continuous in their arguments. These functionals are usually referred to as *vacuum expectation values;* their relation to the field A is somewhat analogous to that between the moments of a probability measure and the measure itself. Their physical interpretation is as a quantitative description of the correlations between field measurements in the vacuum state.

By a straight forward application of Schwartz's nuclear theorem, one can prove that the vacuum expectation values can be uniquely extended to become continuous linear functionals on $\mathcal{D}_{x_1 \ldots x_n}$, the infinitely differentiable functions of compact support on $4n$ space. So extended, these distributions are usually denoted $(\Psi_0, A(x_1) \ldots A(x_n)\Psi_0)$.

By II, the extended distributions are invariant under the mappings $x_1, \ldots, x_n \to \Lambda x_1 + a, \ldots, \Lambda x_n + a$, so they can be written as distributions $F^{(n)}$ of the $n-1$ four-vector variables $\xi_j = x_j - x_{j+1}$, $j = 1, \ldots, n-1$, invariant under homogeneous Lorentz transformations: $\xi_1, \ldots, \xi_{n-1} \to \Lambda\xi_1, \ldots, \Lambda\xi_{n-1}$. III implies

$$\begin{aligned} &F^{(n)}\,(\xi_1, \ldots, \xi_{j-1}, \xi_j, \xi_{j+1}, \ldots, \xi_n) \\ &= F^{(n)}\,(\xi_1, \ldots, \xi_{j-1} + \xi_j, -\xi_j, \xi_j + \xi_{j+1}, \ldots, \xi_n) \end{aligned} \tag{18}$$

for space-like ξ_j. These are customarily called the *local commutativity conditions.*

$F^{(n)}$ obviously satisfies the condition

$$F^{(n)}(\xi_1,\ldots,\xi_{n-1}) = F^{(n)}(-\xi_{n-1},\ldots,-\xi_1) \tag{19}$$

by virtue of (14). These are usually referred to as the *hermiticity conditions.*

The inequality

$$\begin{aligned}&|(\Psi_0, A(f_1)\ldots A(f_{j-1})\,\mathcal{U}(a,1)\,A(f_j)\ldots A(f_n)\Psi_0)|\\ &\leqslant \|A(f_{j-1})^*\ldots A(f_1)^*\Psi_0\|\;\|A(f_j)\ldots A(f_n)\Psi_0\|\end{aligned} \tag{20}$$

yields a boundedness property of $F^{(n)}$ which permits one to conclude that it can be extended to be a continuous linear functional on $\mathcal{S}$ in its jth argument the others being held fixed in $\mathcal{D}$. The Fourier transform in the jth variable vanishes outside the future light cone by virtue of the assumed properties of $\mathcal{U}(a,1)$. This is usually referred to as the *spectrum condition.* Using the standard extension of a Fourier to a Laplace transform one sees that $F^{(n)}$ is the boundary value of a function analytic in its jth variable in a tube, the others being smeared with test functions in $\mathcal{D}$. If it had been assumed that the fields were defined with test functions in $\mathcal{S}$ one would get the result that $F^{(n)}$ is the boundary value of an analytic function when all the variables $\xi_{j-i}\eta_j$ run over the tube $(\eta_j)^2>0$ $\eta_j^0>0$ $j=1,\ldots,n-1$. A significant part of the proof of theorem 3 of the next section is to recover this analyticity in all the variables under the weakened assumption ($\mathcal{D}$).

A further property of $F^{(n)}$ can be derived if one uses the following fact about $\mathcal{U}(a,1)$. The only state invariant under space translations in Ψ_0. One finds

$$\begin{aligned}\lim_{a\to\infty} F^{(n)}(\varphi_1,\ldots,\varphi_{j-1},\{a,1\}\varphi_j,\varphi_{j+1},\ldots,\varphi_{n-1}) = F^{(j)}(\varphi_1\ldots\varphi_{j-1})\times\\ F^{(n-j)}(\varphi_j,\ldots,\varphi_{n-1})\end{aligned} \tag{21}$$

for all space like a. This is sometimes referred to as the *cluster decomposition property.*

Lastly, because

$$\|\sum_k \alpha_k A(\varphi_{k1})\ldots A(\varphi_{kk})\Psi_0\|^2 \geqslant 0$$

for any finite set of complex numbers α_k,

$$\begin{aligned}&\sum_{k,l}\bar{\alpha}_k\alpha_l\int\ldots\int\bar{\varphi}_{kk}(x_k)\ldots\bar{\varphi}_{k1}(x_1)\,\varphi_{l1}(y_1)\ldots\varphi_{ll}(y_l)\\ &\qquad F^{(k+l)}(x_k-x_{k-1},\ldots,x_2-x_1,x_1-y_1,y_1-y_2,\ldots,y_{l-1}-y_l)\\ &\qquad dx_k\ldots dx_1\,dy_1\ldots dy_l\geqslant 0\end{aligned} \tag{22}$$

These are sometimes referred to as the *positive definiteness inequalities.*

It is plausible that, given the vacuum expectation values, one ought to be able in some sense to reconstruct the field and $\mathcal{H}$ and $\mathcal{U}$, at least in the subspace spanned by vectors of the form $P(A(\varphi_1)\ldots)\Psi_0$. Here P is a polynomial in the fields smeared with test functions in $\mathcal{D}$. This linear set of

vectors will be denoted by D_0. Evidently D_0 is a subset of the D of Axiom I. By virtue of II, $\mathcal{U}(a,\Lambda)D_0 \subset D_0$.

THEOREM 1. [5]. *For each $n=0,1,2,\ldots$ let $F^{(n)}$ be a distribution in $\mathcal{D}'$ depending on $(n-1)$ four-vector variables, and invariant under* the *transformations of the homogeneous Lorentz group $\xi_1,\ldots,\xi_{n-1}\to\Lambda\xi_1,\ \ldots,\ \Lambda\xi_{n-1}$. Suppose the $F^{(n)}$ are extendable to $\mathcal{S}$ in each of their arguments the others being held fixed in $\mathcal{D}$. If the $F^{(n)}$ satisfy*

the hermiticity conditions, the spectral conditions, the positive definiteness conditions, the cluster decomposition property,

then there exists a Hilbert space $\mathcal{H}$, a continuous unitary representation of the Lorentz group $\{a,\Lambda\}\to\mathcal{U}(a,\Lambda)$ with energy momentum spectrum in or on the future light cone and unique vacuum Ψ_0, and a hermitean scalar field A satisfying axioms I and II with $D=D_0$ and such that

$$(\Psi_0, A(x_1)\ldots A(x_n)\Psi_0) = F^{(n)}(x_1-x_2,\ldots,x_{n-1}-x_n).$$

This realization is unique up to unitary equivalence.

If, in addition the $F^{(n)}$ satisfy the local commutativity conditions then Axiom III is also satisfied.

This theorem has a more or less trivial extension to the case of a denumerable set of neutral scalar fields. One need only require that I and II are valid for each A in the set with a common dense domain D as both φ and A vary. III is usually extended to be a commutation relation between pairs of fields and their adjoints. There is an $F^{(n)}$ for every word formed from A's; the appropriate generalizations of the conditions on the F's are clear.

The idea that a denumerable set of fields acting in the subspace spanned by D_0 should be regarded as a complete physical theory has recently received strong support as will be explained below.

4. Self-adjoint extensions of hermitean fields

The $A(\varphi)$ for φ real are hermitean when considered as defined on D_0: $(A(\varphi)|_{D_0})^* \supset A(\varphi)|_{D_0}$. It is some interest both as a matter of principle and as a calculational tool to know whether $A(\varphi)|_{D_0}$ possesses self-adjoint extensions. For the free field the best possible result holds

THEOREM 2. *If φ is real and $\in\mathcal{D}$, the free field operator $A(\varphi)$ restricted to $D_0, A(\varphi)|_{D_0}$, is essentially self-adjoint.*

I have so far been unable to extend this result to a general local field. However, the following weaker result can be established.

THEOREM 3. *If φ is real and $\in\mathcal{D}$ and D_0 is dense, then $A(\varphi)|_{D_0}$ has equal defect indices.*

Thus $A(\varphi)|_{D_0}$ has at least one self adjoint extension.

The proof of Theorem 3 is obtained by using a basic symmetry possessed by all local field theories, so-called *PCT symmetry* [6]. PCT symmetry means that there exists (in any theory with a cyclic vacuum) an antiunitary operator, Θ, satisfying

$$\Theta A(\varphi)\Theta^{-1} = A(\hat{\varphi}), \tag{23}$$

where $$\hat{\varphi}(x) = \overline{\varphi(-x)}.$$

PCT symmetry, according to our best present experimental information, holds exactly in Nature (although certain other inversion symmetries, notably space inversion, do not!), and accounts for such phenomena as the exact equality of the mass of proton and anti-proton. For a real, even φ, (23) says that $A(\varphi)|_{D_0}$ commutes with an anti-unitary operator. A standard argument [7] then shows that Θ maps the manifolds of solutions of $(A(\varphi)|_{D_0})^*\Phi = = \pm i\Phi$ onto each other one to one, so $A(\varphi)|_{D_0}$ has equal defect indices.

The standard proof of the PCT theorem uses the assumption that fields are defined for test functions in $\mathcal{S}$. Thus it is essential for the above argument to prove it again with $\mathcal{S}$ replaced by $\mathcal{D}$. The decisive step is again to show under the weaker assumption that the $F^{(n)}$ are boundary values of invariant functions analytic when the variables $\xi_1 - i\eta_1, \ldots, \xi_{n-1} - i\eta_{n-1}$ vary over the tube $\eta_j^2 > 0, \eta_j^0 > 0$, $j = 1, \ldots, n$. Here a theorem proved by M. Zerner [8] is very helpful. Stated for the simplest case of two complex variables, it says: If f is a function of two variables z_1, z_2 defined and continuous when $\operatorname{Im} z_1 = 0$, $\operatorname{Im} z_2 \geqslant 0$ and when $\operatorname{Im} z_2 = 0$, $\operatorname{Im} z_1 \geqslant 0$, and analytic in z_2 for z_1 real, $\operatorname{Im} z_2 > 0$ and also in z_1 for z_2 real, $\operatorname{Im} z_1 > 0$, then there exists a unique function analytic in the product of the upper half planes and continuous on the boundary which coincides with f there.

It should be emphasized that in proving PCT symmetry one does not assume any invariance under space-time inversions. It comes about essentially because 1 and -1 lie in the same component of the complex homogeneous Lorentz group.

There is at the moment no evidence against the conjecture that Theorem 3 could be sharpened to assert the essential selfadjointness of $A(\varphi)|_{D_0}$.

Theorem 3 makes available the spectral resolution for a self-adjoint extension $\hat{A}(\varphi)$ of $A(\varphi)|_{D_0}$ for φ real and $\in \mathcal{D}$:

$$\hat{A}(\varphi) = \int_{-\infty}^{\infty} \lambda dE(\lambda, \varphi). \tag{24}$$

Introducing also the spectral resolution of $U(a, 1)$

$$U(a, 1) = \int e^{-ip.a} dF(p), \tag{25}$$

one can study the quantities

$$(\Psi_0, E(\lambda_1, \varphi_1)\, F(S_1)\, E(\lambda_2, \varphi_2)\, F(S_2) \ldots E(\lambda_n, \varphi_n)\Psi_0), \tag{26}$$

where the S_j are measurable sets of the energy momentum spectrum. (26) when integrated appropriately yields the vacuum expectation value $F^{(n)}$.

Concluding remarks

The preceding results, while perhaps satisfying to the technicians struggling with quantum field theory, deal mainly with scaffolding. One would like to get some idea what the building might look like if it is ever completed

and doesn't fall down. Could one, for example, simplify life mathematically by assuming that $\mathcal{U}$ contains the identity representation and one other irreducible. Perhaps one could then find non-tuvial examples of fields satisfying I, II, III. This is settled negatively by

THEOREM 4. *The spectrum of* $\mathcal{U}(a,1)$ *is an additive set in a theory in which* Ψ_0 *is cyclic.*

Much deeper than Theorem 4 are the recent results of D. Ruelle [9] which show that if any irreducible representation say $[m,j]$ occurs in $\mathcal{U}$, then $\bigoplus_{n=0}^{\infty} \otimes^s [m,j]^n$ occurs in $\mathcal{U}$. Here $\otimes^s [m,j]^n$ is the symmetrized tensor product of $[m,j]$ with itself n times. More in detail what Ruelle shows is that, implicit in any theory of a denumerable set of fields satisfying I, II, III and having a cyclic vacuum state is an essentially uniquely determined collision theory for particles characterized by the masses and spins of the irreducible representations contained in $\mathcal{U}$. The only additional assumption necessary to make the theory physically satisfactory is (Ruelle's Axiom IV) that these collision states span the whole Hilbert space. Coupled with Borchers recent result [10] that a denumerable set of fields satisfying I, II, III which has a cyclic vacuum is necessarily irreducible, this is strong evidence that a theory of fields satisfying I, II, III with a cyclic vacuum is a reasonable object of study. However, it also makes plausible that there is no interesting way of simplifying the problem mathematically; every non-trivial solution must have at least the complexity one associates with non-trivial collision processes [11].

REFERENCES

[1]. Although the remarks of this introduction will be familiar to all who have followed the development of quantum field theory, they seem appropriate here. A more complete discussion with references is contained in: *Problèmes Mathématiques de la Théorie Quantique des Champs*, pp. 1–38. Centre National de la Recherche Scientifiques Paris, 1959.

[2]. A proof of the uniqueness of the free field follows from the results of A. S. Wightman and S. S. Schweber: Configuration space methods in relativistic quantum field theory. *Phys. Rev.*, 98 (1955), 812.

[3]. The axioms in the form used here are due to Gårding and Wightman; see reference [1]. Here, for simplicity, only scalar fields are discussed.

[4]. WIGHTMAN, A. S., Quantum field theory in terms of vacuum expectation values. *Phys. Rev.*, 101 (1956), 860.

[5]. Theorem 1 has its origin in [4]. However, as stated here it has been sharpened to incorporate the discovery (due to K. Hepp, R. Jost, D. Ruelle, and O. Steinmann: Necessary restriction on Wightman functions. *Helv. Phys. Acta*, 34 (1961), 542; and H. J. Borchers: On structure of the algebra of field operators. *Nuovo Cimento* 24 (1962), 214) that the cluster decomposition property guarantees the uniqueness of the vacuum.

[6]. JOST, R., Eine Bemerkung zum CTP Theorem. *Helv. Phys. Acta*, 30 (1957), 409.

[7]. STONE, M. H., *Linear Transformations in Hilbert Space*, 357–365.

[8]. ZERNER, M., Private communication and Séminaire de Physique Mathématique de la Faculté des Sciences de Marseille. There is an earlier as yet unpublished proof by Malgrange, as well as an independent unpublished proof by Aronszajn.

[9]. RUELLE, D., On the asymptotic condition in quantum field theory. *Helv. Phys. Acta*, 35 (1962) 147.

[10]. BORCHERS, H. J., reference 5. See also Appendix of reference 9.

[11]. Excepting the interesting possibility of analogues of relativistic field theory in a world with a finite number of points. See Hans Joos: Group theoretical models of local field theories. *J. Math. Phys.* (in press).

INDEX

Adams, J. F., 435
Agmon, S., 301
Ahlfors, L. V., 3
Andreotti, A., 306
Atiyah, M. F., 442
Auslander, M., 230

Baily, W. L. Jr., 309
Berger, M., 447
Bing, R. H., 457
Borel, A., 10

Carleson, L., 314
Cassels, J. W. S., 234
Choquet, G., 317
Church, A., 23
Cohen, P. J., 331

Dold, A., 459
Dwork, B., 247
Dynkin, E. B., 36

Eckmann, B., 59
Ehrenpreis, L., 337

Floyd, E. E., 462

Ganea, T., 467
Gel'fand, I. M., 74
Grad, H., 560
Grauert, H., 86

Henrici, P., 102
Hironaka, H., 507
Hunt, G. A., 531
Hörmander, L., 339

Igusa, J., 522
Itô, K., 536

Jenkins, J. A., 347

Kahane, J.-P., 114
Katětov, M., 473
Kervaire, M. A., 480
Kneser, M., 260
Kolmogorov, A. N., 351
Kostrikin, A. I., 264
Kuranishi, M., 357

Leray, J., 360
Linnik, Ju. V., 270
Łoś, J., 225

Malliavin, P., 368
Milnor, J., 132
Moser, J., 584
Mumford, D., 526

Nachbin, L., 379
Narasimhan, R., 385
Newman, M. H. A., 139
Nirenberg, L., 147

Pjateckiĭ-Šapiro, I. I., 389
Pliś, A., 397
Poénaru, V., 481

Šafarevič, I. R., 163
Selberg, A., 177
Serre, J.-P., 190
Shilov, G. E., 403
Sinaĭ, Ja. G., 540
Smale, S., 490
Smirnov, Ju., 497
Stampacchia, G., 405
Stein, E. M., 414
Suzuki, M., 285
Szőkefalvi-Nagy, B., 421

Tate, J., 288
Thompson, J. G., 296
Tits, J., 197

Wermer, J., 426
Whitehead, G. W., 502
Wightman, A. S., 587